DOCUMENT RESUME

ED 434 928 TM 030 180

| | |
|---|---|
| AUTHOR | Lane, Suzanne; Ventrice, James; Cerrillo, Tracy L.; Parke, Carol S.; Stone, Clement A. |
| TITLE | Impact of the Maryland School Performance Assessment Program (MSPAP): Evidence from the Principal, Teacher and Student Questionnaires (Reading, Writing, and Science). |
| SPONS AGENCY | Department of Education, Washington, DC. |
| PUB DATE | 1999-04-00 |
| NOTE | 54p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Montreal, Quebec, Canada, April 19-23, 1999). |
| PUB TYPE | Reports - Research (143) -- Speeches/Meeting Papers (150) |
| EDRS PRICE | MF01/PC03 Plus Postage. |
| DESCRIPTORS | Educational Assessment; Educational Change; *Educational Objectives; Educational Practices; Elementary Education; *Language Arts; Middle Schools; *Performance Based Assessment; Principals; Reading; Science Education; State Programs; *Student Attitudes; Student Surveys; *Teacher Attitudes; Teacher Surveys; Teachers; Testing Programs; Writing (Composition) |
| IDENTIFIERS | *Maryland School Performance Assessment Program; Reform Efforts |

ABSTRACT

This study examines the underlying structure of the teacher language arts and science questionnaires from a project studying the impact of the Maryland School Performance Assessment Program (MSPAP). The dimensions of the questionnaire concerning practices and attitudes related to the MSPAP were studied, as well as the extent to which the on-grade and off-grade teachers differed on these dimensions. The extent to which principals and students differed from teachers with respect to a subset of the dimensions was also examined. The final samples for language arts included the teachers, principals, and some students from 59 elementary and 31 middle schools. For the science questionnaire, the final sample contained 103 elementary and 58 middle schools. Overall, evidence suggests that the MSPAP is having an impact on classroom instruction to some extent. The majority of the language arts and science teachers indicated that they have made changes in their classroom activities to reflect the Maryland learning outcomes and the MSPAP. According to the teachers, the impact has been similar between the on- and off-grade levels. When differences occurred, they were primarily between elementary and middle school teachers, with elementary school teachers tending to respond more favorably than middle school teachers. Most teachers indicated that the MSPAP is a useful tool for improving instruction, but the majority also indicated that they oppose using the MSPAP for identifying schools for rewards or recognition. (Contains 29 tables and 21 references.) (SLD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED 434 928

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Suzanne Lane
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

# Impact of the Maryland School Performance Assessment Program (MSPAP): Evidence From the Principal, Teacher and Student Questionnaires (Reading, Writing, and Science)

by

Suzanne Lane, James Ventrice, Tracy L. Cerrillo, Carol S. Parke, and Clement A. Stone
University of Pittsburgh

Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, April 1999. Preparation of this paper was supported by a grant from the U.S. Department of Education, Assessment Development and Evaluation Grants Program (CFDA 84.279-A), for the Maryland Assessment System Project. Our deepest appreciation is extended to the teachers, principals, and students in Maryland for their invaluable time and effort spent on this project.

TM030180

BEST COPY AVAILABLE

Impact of the Maryland School Performance Assessment Program (MSPAP):
Evidence form the Principal, Teacher and Student Questionnaires
(Reading, Writing, and Science)

A number of states are implementing statewide assessment programs that depend heavily on performance-based assessments (e.g., Kentucky, Maryland). These assessments are considered critical tools in the educational reform movement (Linn, 1993) and are being used for high-stakes purposes such as holding schools accountable to state standards. A prevailing assumption underlying performance-based assessments is that they serve as motivators in improving student achievement and learning, and that they encourage instructional strategies and techniques that foster reasoning, problem solving, and communication (Frederiksen & Collins, 1989; National Council on Education Standards and Testing, 1992).

Given these high expectations for performance-based assessments, the consequences of the uses and interpretations of the assessments need to be addressed, including both negative and positive consequences, intended and plausible unintended consequences (Messick, 1989, 1992; Cronbach, 1988; Koretz, Barron, Mitchell, & Stecher, 1996; Linn, Baker, & Dunbar, 1991). As stated by Linn (1994), "If the argument that validation should include an evaluation of the consequences of the uses and interpretations of assessment results is accepted, then it is not sufficient to provide evidence that the assessments are measuring the intended constructs. Evidence is also needed that the uses and interpretations are contributing to enhanced student achievement and at the same time, not producing unintended negative outcomes (p. 8)."

Researchers are beginning to examine the impact of assessment programs by using various methods such as surveys of principals and teachers (e.g., Koretz, Barron, Mitchell, & Stecher, 1996; Pomplum, 1997) and focus groups (e.g., Chudowsky & Behuniak, 1997). The research program, for which this set of studies is a part of, attempts to supplement self-report data with more direct evidence, evidence obtained through the analyses of classroom instruction and assessment activities. Other studies have also included the use of classroom activities but to a limited extent (e.g., McDonnell & Choisser, 1997). In addition to conducting interviews, McDonnell and Choisser (1997) collected classroom artifacts from 23 teachers in both Kentucky and North Carolina over a two-week period.

The purpose of this research program is to examine the impacts of the Maryland State Performance Assessment Program (MSPAP) and the Maryland Learning Outcomes (MLO's) on classroom instruction and assessment practices, student learning, professional development, and students, teachers, and principals beliefs about MSPAP. MSPAP is a performance assessment program for grades 3, 5, and 8

designed to measure school performance and provide information for school accountability and improvement so as to ensure quality education (Maryland State Board of Education, 1995). MSPAP was implemented in the earlier 1990's to assess student achievement and school performance with respect to the Maryland Learning Outcomes. On MSPAP, students develop written responses to interdisciplinary tasks that require the application of skills and knowledge to real life problems, and is intended to promote performance-based instruction and classroom assessments.

The research questions for the project are: (1) What are the impacts of MSPAP on classroom instruction and assessment practices; student learning; professional development activities; school-based decision-making; and student, teacher and principal beliefs and attitudes? (2) How do the impacts vary by content area (mathematics, reading, writing, science, social studies), grade level (on-grades: 3, 5, 8 and off-grades: 2, 4, 7), and school characteristics (percent free or reduced lunch and MSPAP performance)? This study described herein is limited to examining the impact of MSPAP for the 1996-97 instructional year for the reading and writing content areas and for the 1997-98 instructional year for the science content area in elementary and middle schools in Maryland.

Researchers have argued that rigorous evidence for the consequences of an assessment has yet to be obtained (Kane, Khattri, Reeve, & Adamson, 1997; Mehrens, 1998). Mehrens (1998) further states that causative inferences cannot be drawn from the evidence collected thus far. The evidence that is presented in the present set of studies is also limited, in that causative inferences can not be drawn. However, as suggested by Mehrens, it may be reasonable to determine whether the evidence "suggests" positive and/or negative consequences. It is the goal of this project to provide additional evidence suggesting the impacts that assessment programs have on instruction and learning.

**Objectives of the Symposium Papers**

The first paper examines the underlying structure of the language arts teacher questionnaire (reading and writing) administered to teachers in the 1996-97 instructional year and the science teacher questionnaire administered to teachers in the 1997-98 year. In general, confirmatory factor analyses (CFA; Joreskgog, 1969; Joreskog & Sorbom, 1979) were conducted to examine the existence of the hypothesized dimensions related to teachers' support for MSPAP, teachers' emphasis on learning outcomes and on reform-oriented problem types in instruction, teachers' change in emphasis on learning outcomes and on reform-oriented problem types in instruction, MSPAP's impact on instruction, and MSPAP-related professional development activities. Analysis of variances were then conducted to determine the extent to which teachers of the on- and off-grade levels differed on these dimensions. Similar analyses were conducted on the principal and student questionnaires to determine the extent to

which elementary and middle school principals as well as on- and off-grade students differed on a subset of the dimensions.

The second paper presents the results of the analyses of the language arts instruction, assessment and test preparation activities. The goal of this study was to determine the extent to which the classroom activities were aligned to the Maryland Learning Outcomes and MSPAP.

The third paper examines the relationship among science school performance gains on MSPAP, the percentage of students who were receiving a free or reduced lunch in the schools, which served as a proxy for socioeconomic level, and the impact MSPAP has had on instruction. More specifically, a growth model analysis (c.f., Meredith & Tisak, 1990; McArdle & Epstein, 1987; Muthen, 1994) was conducted to examine the relationship among changes in MSPAP school performance in science, percent free or reduced lunch, classroom instruction and assessment practices, MSPAP's impact on classroom instruction and assessment practices, and student motivation.

## Methodology

### School Sample

School Sample for Language Arts. A stratified random sampling procedure was used to select the schools, with the strata being defined by three levels of each of the following: (a) percent free or reduced lunch according to the 1994-95 classification and (b) MSPAP performance gains (MSDE's 1993-95 change index). Schools were classified into one of the nine cells based on their rankings in the distributions for these two variables. Eight elementary schools from each of the nine cells were sampled and four middle schools from each of the nine cells were sampled. A total of 72 elementary and 36 middle schools were selected to participate in the study with alternate schools identified as potential replacements for schools who chose not to participate. A larger number of elementary schools were selected because, compared to the middle schools, they have fewer teachers per grade.

The final sample consisted of 59 elementary and 31 middle schools, with a total of 90 schools. Thus, the school participation rate was 82% for elementary schools and 86% for middle schools. There were approximately equal numbers of schools within each of the nine classification cells. Of the 59 elementary schools, 42 were from the initial 72 that were sampled, and of the 31 middle schools, 22 were from the initial 36 that were sampled. The remaining schools were from the list of alternate schools for each cell. This represents schools from 19 systems/counties in Maryland. It should be noted that, because schools were unable to be contacted until January 1997 regarding their participation in the study, the sample size for the 1996-97 instructional year was reduced.

School Sample for Science. The same stratified random sampling procedure was used to select the schools for the science data with the strata being defined by three levels of each of the following: (a) percent free or reduced lunch according to the 1994-95 classification and (b) MSPAP performance gains (MSDE's 1993-95 change index). Prior to selecting schools for the science area, however, those schools that participated in the data collection for language arts in the 1996-97 year were excluded. Schools were classified into one of the nine cells based on their rankings in the distributions for these two variables. Fourteen elementary schools from each of the nine cells were sampled and seven middle schools from each of the nine cells were sampled. A total of 126 elementary and 63 middle schools were selected to participate in the study with alternate schools identified as potential replacements for schools who chose not to participate. A larger number of elementary schools were selected because, compared to the middle schools, they have fewer teachers per grade. A larger sample size for the 1997-98 year was used in order to maximize the number of schools participating in the project.

The final sample consisted of 103 elementary and 58 middle schools, with a total of 161 schools. Thus, the school participation rate was 82% for elementary schools and 92% for middle schools. There were approximately equal numbers of schools within each of the nine classification cells. Of the 103 elementary schools, 87 were from the initial 126 that were sampled, and of the 58 middle schools, 44 were from the initial 63 that were sampled. The remaining schools were from the list of alternate schools for each cell. This represents schools from 22 systems/counties in Maryland. In summary, across the two years, a total of 251 schools participated in the study.

**Instruments**

To triangulate on the consequences of MSPAP, multiple measures were used. The data sources used for this set of studies were questionnaires and samples of classroom instruction, assessment, and test preparation materials. Questionnaires were developed for principals, teachers and students. The questionnaire for principals was the same for both elementary and middle school principals. Separate Language Arts and Science questionnaires were developed for 2nd, 3rd, 4th, 5th, 7th, and 8th grade teachers in those respective areas. The teacher questionnaires did not vary substantially across on- and off-grades (i.e., tested and not tested grades, respectively). Language Arts and Science questionnaires were developed for students in 4th, 5th, 7th, and 8th grades. The questionnaires for the 4th and 7th grade (i.e., off-grade) students contained a MSPAP public release task so that the students could examine the task prior to responding to questions pertaining to MSPAP-like tasks.

The questionnaires consisted of both Likert and constructed response items. Some of the Likert items were in the form of questions, and others were statements. In general, a four-point scale was used

for the Likert items. To triangulate on the evidence, students, teachers, and principals responded to similar questions for areas in which it was deemed appropriate. In general, the teacher questionnaires contained items pertaining to the following areas: Familiarity with MSPAP, support for MSPAP, beliefs about MSPSP, the nature of instruction and classroom assessments, MSPAP's impact on instruction and classroom assessments, the nature of professional development activities, and MSPAP's impact on professional development activities. The principal and student questionnaires included items for areas that were deemed appropriate. Some of the ideas for questions pertaining to the support for MSPAP and the beliefs about MSPAP were based on a previous study examining the consequential evidence of state assessments (Koretz, Mitchell, Baron, & Keith, 1996).

The language arts questionnaires were piloted in the spring of 1995 in schools in Maryland and were reviewed by Maryland Language Arts teachers. The science questionnaires were reviewed by Maryland science teachers in the fall of 1996.

**Data Collection**

Data Collection for Language Arts (1996-97 year). Teachers and principals were asked to complete their respective questionnaires during February 1997. Students were administered the student questionnaire within the two weeks following the administration of MSPAP, that is, in either the 3rd or 4th week of May 1997.

Data Collection for Science (1997-98). Teachers and principals were asked to complete their respective questionnaires during February 1998. Students were administered the student questionnaire within the two weeks following the administration of MSPAP, that is, in either the 3rd or 4th week of May 1998.

**Questionnaire Return Rate**

Principal and Teacher Questionnaire. For the 1996-97 year, a total of 86 of the 90 principals completed the principal questionnaire, resulting in a total response rate of 96%. A total of 505 2nd, 3rd, 4th, 5th, 7th, and 8th grade teachers out of 593 completed the language arts teacher questionnaires, with a total response rate of 85%. Table 1 indicates the return rates for principals and language arts teachers in elementary and middle schools for the 1996-97 year. The number of language arts teachers in each grade level that completed the questionnaires are 69 2nd grade teachers, 90 3rd grade teachers, 70 4th grade teachers, 92 5th grade teachers, 73 7th grade teachers, and 111 8th grade teachers.

For the 1997-98 year, a total of 147 of the 161 principals completed the principal questionnaire, resulting in a total response rate of 91%. A total of 682 2nd, 3rd, 4th, 5th, 7th, and 8th grade science

teachers out of 917 completed the teacher questionnaires, with a total response rate of 75%. The number of science teacher in each grade that completed the questionnaires are 130 2nd grade teachers, 130 3rd grade teachers, 101 4th grade teachers, 107 5th grade teachers, 97 7th grade teachers, and 130 8th grade teachers.

Table 1
Principal and Teacher Questionnaire Return Rate

| | Completed Questionnaire | | Return |
|---|---|---|---|
| | Actual | Proposed | Rate |
| 1996-97 Year | | | |
| Principals | | | |
| Total | 86 | 90 | 96% |
| Elementary | 58 | 59 | 98% |
| Middle | 28 | 31 | 90% |
| Lang. Arts Teachers | | | |
| Total | 505 | 593 | 85% |
| Elementary | 321 | 372 | 86% |
| Middle | 184 | 221 | 83% |
| 1997-98 Year | | | |
| Principals | | | |
| Total | 147 | 161 | 91% |
| Elementary | 95 | 103 | 92% |
| Middle | 52 | 58 | 90% |
| Science Teachers | | | |
| Total | 683 | 927 | 74% |
| Elementary | 468 | 621 | 78% |
| Middle | 215 | 306 | 70% |

Student Questionnaire. For the 1996-97 year, each of the 4th, 5th, 7th, and 8th grade language arts teachers participating in the study were asked to administer the student questionnaire to one of their classes. Overall, 115 of the 156 elementary classes (4th and 5th grades) that were identified for the administration of the language arts student questionnaires actually administered the questionnaires, resulting in a return rate of 74%. In the middle school classes (7th and 8th grades), 118 of the 170 identified classes administered the language arts student questionnaires (69%). Table 2 indicates the number of students and classes in 4th, 5th, 7th, and 8th grades who completed the language arts questionnaires. It should be noted that each of the questionnaires were divided into 3 forms and a student only received one form. The forms were randomly distributed within each of the participating classrooms. This sampling design was used to reduce the amount of time taken away from instruction.

For the 1997-98 year, each of the 4th, 5th, 7th, and 8th grade science teachers participating in the study were asked to administer the student questionnaire to one of their classes. Overall, 203 of the 263 elementary classes (4th and 5th grades) that were identified for the administration of the science student questionnaires actually administered the questionnaires, resulting in a return rate of 77%. In the middle school classes (7th and 8th grades), 218 of the 299 identified classes administered the science student questionnaires (73%).

Table 2
Student Questionnaire Return Rate

| | # of students | # of classes | Average # of students per form |
|---|---|---|---|
| Language Arts (1996-97) | | | |
| 4th | 1197 | 54 | 399 |
| 5th | 1291 | 61 | 430 |
| 7th | 1201 | 53 | 400 |
| 8th | 1358 | 65 | 453 |
| Science (1997-98) | | | |
| 4th | 2269 | 101 | 756 |
| 5th | 2320 | 102 | 773 |
| 7th | 2490 | 102 | 832 |
| 8th | 2721 | 116 | 907 |

**Analyses**

Confirmatory factor analyses (CFA; Joreskog, 1969; Joreskog & Sorbom, 1979) were conducted to examine the underlying structure of the language arts and science teacher questionnaires. To determine whether on- and off-grade teachers differed in their responses to the dimensions identified by the confirmatory factor analyses an analysis of variance was conducted for each dimension. In addition, similar analyses (ANOVA's) were conducted for the dimensions in the principal and student questionnaires.

## Results and Discussion

### Language Arts Teacher Questionnaire

**Confirmatory Factory Analyses for the Language Arts Teacher Questionnaire**

The language arts teacher questionnaire was designed to provide information about ten dimensions. They included teachers' familiarity with MSPAP, teachers' support for MSPAP, teachers' instruction

and assessment practices in reading, teachers' instruction and assessment practices in writing, teacher's use of reform-oriented problem types, change in teachers' instruction and assessment practices in reading, change in teachers' instruction and assessment practices in writing, change in teacher's use of reform-oriented problem types, MSPAP's impact on instruction and classroom assessment, and professional development support for teachers. Subsets of items were related to the following 17 areas (i.e., measures) to reflect the ten dimensions:

(1) MSPAP Familiarity – General (teachers' general familiarity with MSPAP),

(2) MSPAP Familiarity – Results (teachers' familiarity with MSPAP results),

(3) Support MSPAP – General (teachers' general support for MSPAP),

(4) Support MSPAP – Instruction (teachers' support for MSPAP for instructional purposes),

(5) Current Reading Instruction/Assessment- Purposes of Reading,

(6) Current Reading Instruction/Assessment- Stances of Reading,

(7) Current Writing Instruction/Assessment- Purposes for Writing,

(8) Current Writing Instruction/Assessment- Process of Writing,

(9) Current Problem Type (emphasis on reform-oriented problem types in instruction and assessment),

(10) Change Reading Instruction/Assessment- Change in Purposes of Reading (change in emphasis for Purposes of Reading),

(11) Change Reading Instruction/Assessment- Change in Stances of Reading (change in emphasis for Stances of Reading),

(12) Change Writing Instruction/Assessment- Change in Purposes for Writing (change in emphasis for Purposes for Writing),

(13) Change Writing Instruction/Assessment- Change in Process of Writing (change in emphasis for Process of Writing),

(14) Change Problem Type (change in emphasis on reform-oriented problem types in instruction and assessment),

(15) MSPAP's Impact (MSPAP's impact on instruction and assessment),

(16) Professional Development Support - MSPAP (professional development activities related to MSPAP), and

(17) Professional Development Support- Amount (amount of professional development activities).

The measures for Current Reading Instruction and Assessment, Current Writing Instruction and Assessment, Change in Reading Instruction and Assessment, and Change in Writing Instruction and Assessment reflect the Maryland Learning Outcomes (MLO's) in reading and writing.

Teacher mean scores were obtained for each of these seventeen subsets of items (i.e., measures). The majority of the items on the questionnaire had a four-point Likert scale. For those items that had more than a four-point scale, the responses were recoded to a four-point scale. Teacher data were excluded for those cases in which teachers had left blank more than 25% of the items on any one of the seventeen subsets of items. Figure 1 provides the final set of items for each of the measures and the hypothesized dimension underlying each of the measures. Coefficient alpha reliability estimates for these 17 measures for both on- and off-grade data sets ranged from .68 to .95.

Figure 1
Hypothesized Dimensions, Measures, and Language Arts Teacher Questionnaire Items

| Dimension/Measure | Teacher Questionnaire Item |
|---|---|
| MSPAP Familiarity | |
| General | To what extent are you familiar with each of the following? |
| | purpose of MSPAP* |
| | format of MSPAP tasks* |
| | content of skills assessed by MSPAP* |
| | how to prepare students for MSPAP* |
| Results | how to interpret and use MSPAP results to improve classroom instruction and assessment* |
| | how to explain MSPAP results to students and/or parents* |
| Support MSPAP | |
| General | To what extent do you support or oppose MSPAP?* |
| | To what extent has your support or opposition of MSPAP changed over the last few years?* |
| | To what extent do you support or oppose the reporting of MSPAP results?* |
| | To what extent do you support or oppose holding schools accountable for meeting the performance standards on MSPAP?* |
| Instruction | MSPAP is a useful tool for helping me make positive changes in my instruction.* |
| | Results of MSPAP provide useful information for making inferences about school improvement.* |
| Current Reading Instruction/Assessment | |
| Purposes of Reading | About how much emphasis have you placed on each of the following purposes of reading in your classroom activities this year (1996-97)? |
| | reading for a literary experience** |
| | reading to be informed** |
| | reading to perform a task** |
| Stances of Reading | About how much emphasis have you placed on each of the following four stances of reading in your classroom activities this year (1996-97)? |
| | construct a global understanding of text |
| | develop an interpretation of text / modifying understanding |
| | give a personal response to text |
| | take a critical stance |
| Current Writing Instruction/Assessment | |
| Purposes for Writing | About how much emphasis have you placed on each of the following purposes of writing in your classroom activities this year (1996-97)? |
| | writing to inform** |
| | writing to persuade** |
| | writing to express personal ideas** |


BEST COPY AVAILABLE

Figure 1
Hypothesized Dimensions, Measures, and Language Arts Teacher Questionnaire Items

| | |
|---|---|
| Process of Writing | About how much emphasis have you placed on each of the following four stages of the writing process in your classroom activities this year (1996-97)? |
| | prewriting** |
| | drafting** |
| | revising a draft** |
| | proofreading** |
| Current Problem Type | |
| Problem Type | About how often have you used each of the following instructional activities in your language arts classroom this year (1996-97)? |
| | writing essays or several paragraphs** |
| | projects/extended activities** |
| | journal entries** |
| | oral presentations** |
| | group activities** |
| Change in Reading Instruction/Assessment | |
| Purposes of Reading | How has the emphasis on the purposes of reading changed from 1992-93 to 1996-97? |
| | reading for a literary experience |
| | reading to be informed |
| | reading to perform a task |
| Stances of Reading | How has the emphasis on the stances of reading changed from 1992-93 to 1996-97? |
| | construct a global understanding of text |
| | develop an interpretation of text / modifying understanding |
| | give a personal response to text |
| | take a critical stance |
| Change in Writing Instruction/Assessment | |
| Purposes for Writing | How has the emphasis on the purposes of writing changed from 1992-93 to 1996-97? |
| | writing to inform |
| | writing to persuade |
| | writing to express personal ideas |
| Process of Writing | How has the emphasis on the four stages of the writing process changed from 1992-93 to 1996-97? |
| | prewriting |
| | drafting |
| | revising a draft |
| | proofreading |

BEST COPY AVAILABLE

Figure 1
Hypothesized Dimensions, Measures, and Language Arts Teacher Questionnaire Items

| | |
|---|---|
| Change in Problem Type | |
| Problem Type | How has the emphasis for each problem type used in your language arts classroom changed from 1992-93 to 1996-97? |
| | writing essays or several paragraphs |
| | projects/extended activities |
| | journal entries |
| | oral presentations |
| | group activities |
| MSPAP Impact | |
| | About how much has MSPAP influenced your language arts classroom activities this year (1996-97)? |
| | To what extent has MSPAP influenced you to make positive changes in your language arts instruction?* |
| | To what extent has MSPAP influenced you to make positive changes in your language arts assessment?* |
| | About how often do you ask your students solve reading and writing tasks similar to those on MSPAP? |
| | To what extent have you focused on the following strategies in preparing your students for MSPAP? |
| | increasing the use of MSPAP-like tasks in regular instruction* |
| | increasing the match between the content of instruction and the content of MSPAP* |
| Professional Development Support | |
| Focus on MSPAP | To what extent did staff development activities address the following? |
| | purpose of MSPAP |
| | format of MSPAP |
| | content and skills assessed by MSPAP |
| | how to prepare students for MSPAP |
| | how to interpret and use MSPAP results to improve instruction |
| | how to explain MSPAP results to students/parents |
| Amount of Support | To what extent have you had the necessary support and/or resources to enable you to make instructional and assessment changes to better reflect what is expected of students in MSPAP and the Maryland Learning Outcomes? |
| | Support for instruction* |
| | Support for assessment* |

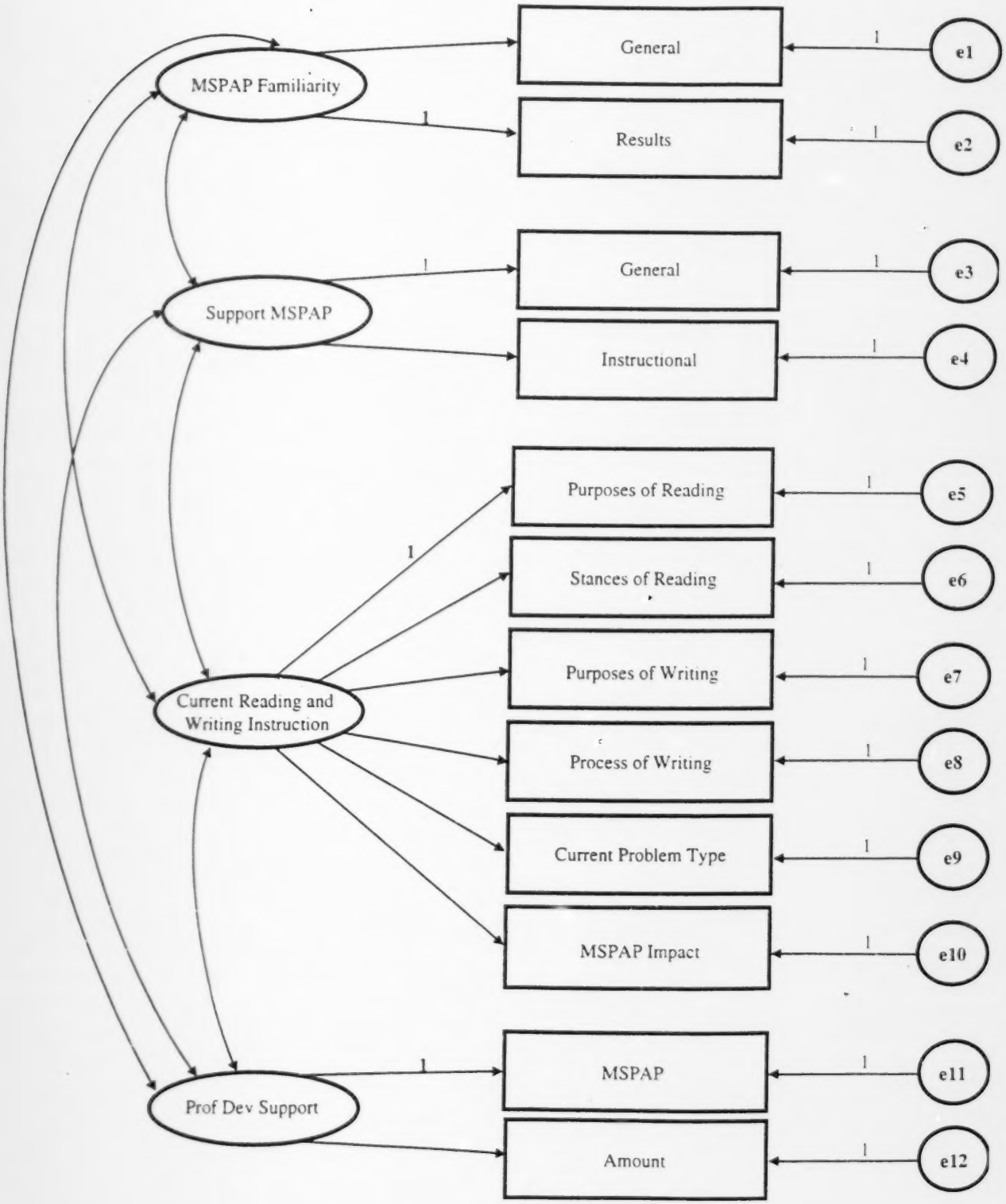
Note: The * items indicate that principals were administered similar items, but from the principal's perspective. The ** items indicate that students were administered similar items, but from the student's perspective.

Two sets of analyses were conducted. For each set of analyses, maximum likelihood estimation procedures were used to estimate three hierarchical models using AMOS (Arbuckle, 1997). The first set excluded the five teacher mean scores related to change in classroom instruction and assessment practices: Change in Purposes of Reading, Change in Stances of Reading, Change in Purposes for Writing, Change in Process of Writing, and Change in Problem Type. The second set of analyses excluded the five teacher mean scores related to current instruction and assessment practices: Purposes of Reading, Stances of Reading, Purposes for Writing, Process of Writing, and Problem Type. Teachers answered the questions with respect to instructional change only if they were teachers in Maryland since the 1992-93 school year. Thus, the second set of analyses is based on a smaller sample size than the first set of analyses.

For the analyses excluding the instructional change measures, the first model that was estimated provided a test for the hypothesis that one factor accounted for the interrelations among the teacher mean scores for twelve measures. The second model that was estimated provided a test for the hypothesis that four factors accounted for the interrelationships as specified in Figure 2. The third model that was estimated, the hypothesized model, provided a test for the hypothesis that seven factors accounted for the interrelationships as specified in Figure 3. For the analyses excluding the current instructional measures, similar models were estimated as shown in Figures 4 and 5. In these analyses the instructional change measures replaced the current instructional measures.

The analyses were conducted for the on-grade levels (3, 5, 8) combined and the off-grade levels (2, 4, 7) combined. The sample sizes for the analyses excluding the instructional change measures were 247 for the on-grade and 176 for the off-grade. The sample sizes for the analyses excluding the current instructional measures were 185 for the on-grade and 126 for the off-grade.

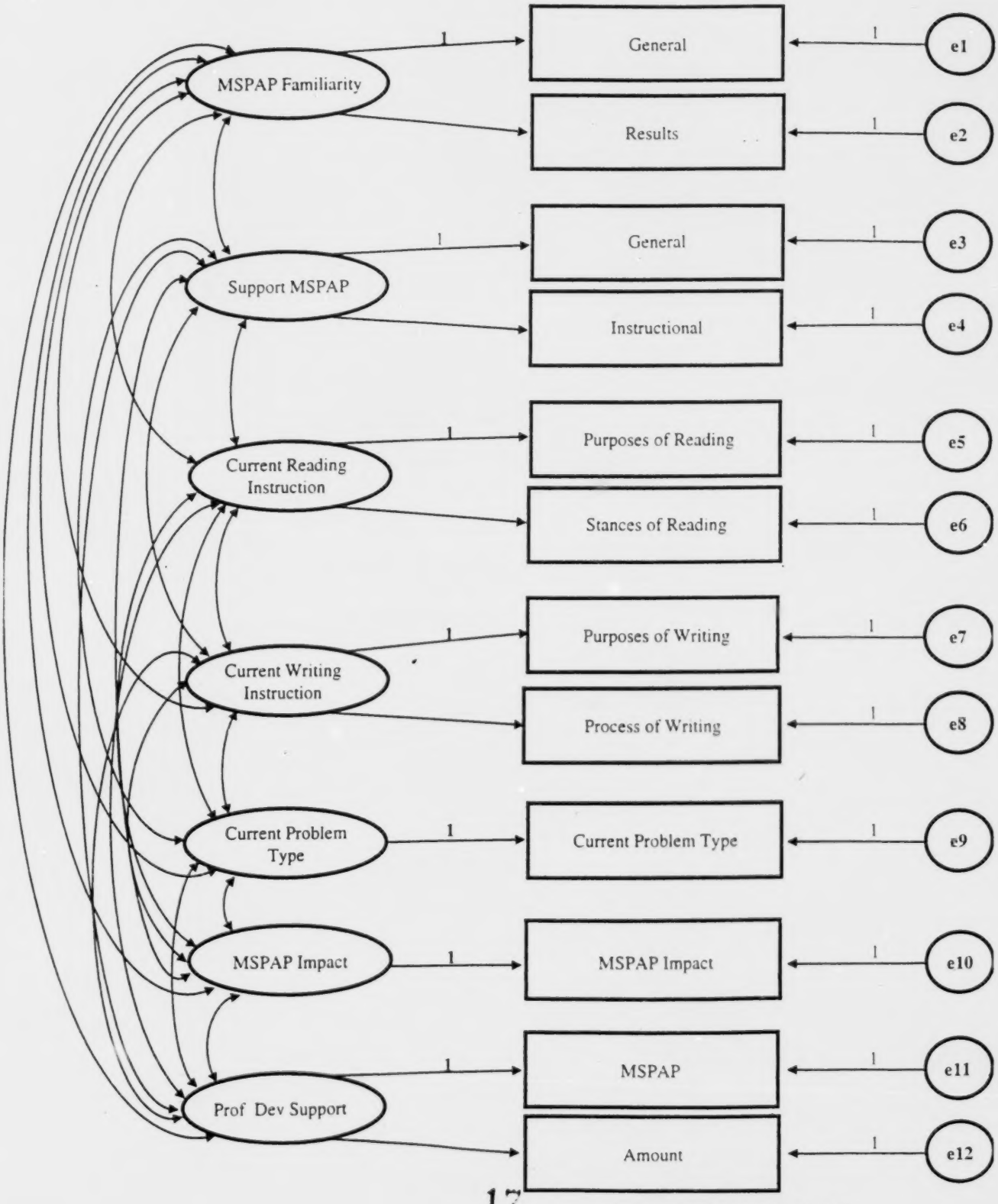
**Figure 2. Four-Factor Model Excluding Change Measures**






BEST COPY AVAILABLE

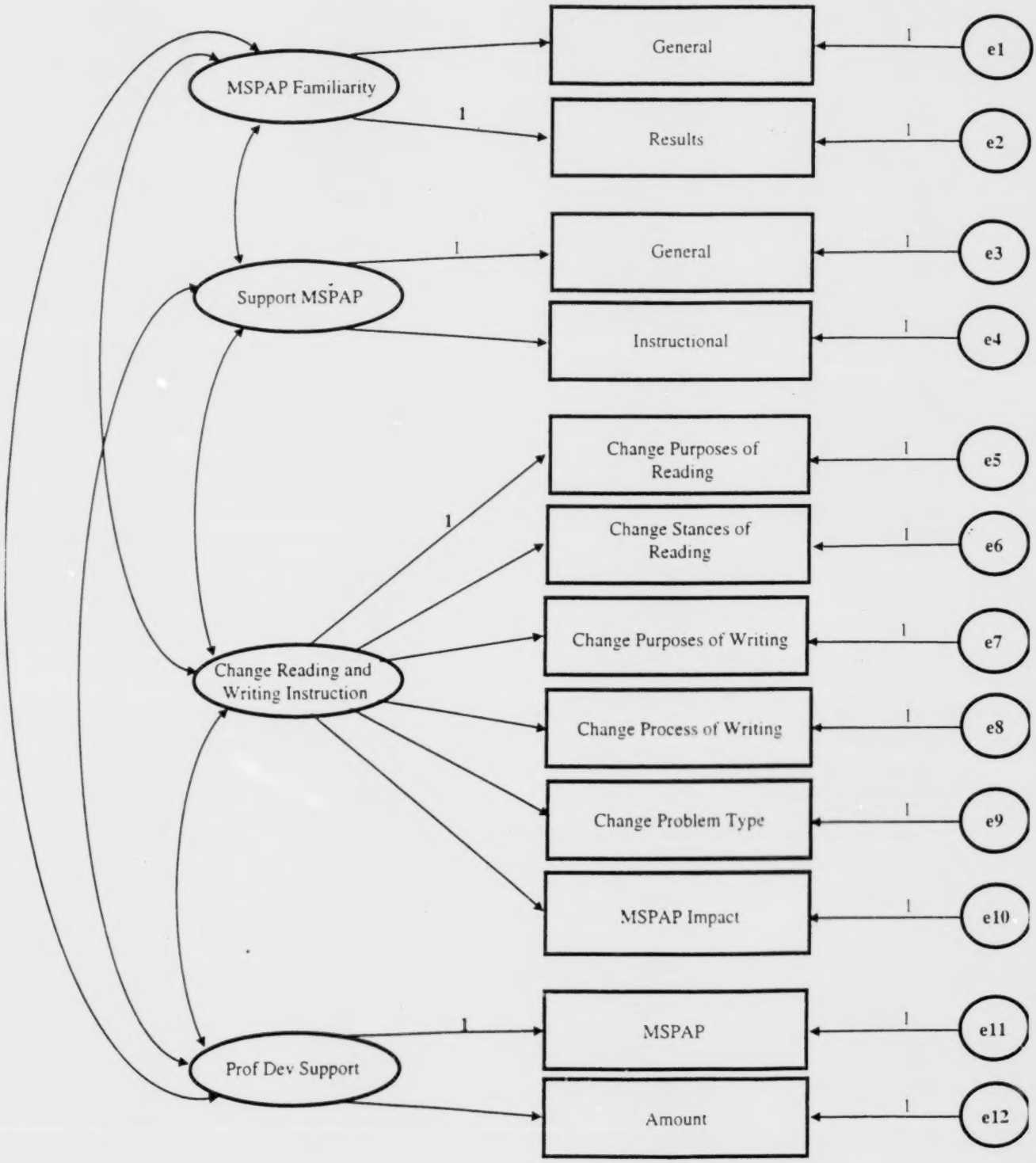
**Figure 3. Seven-Factor Model Excluding Change Instruction Measures**






BEST COPY AVAILABLE

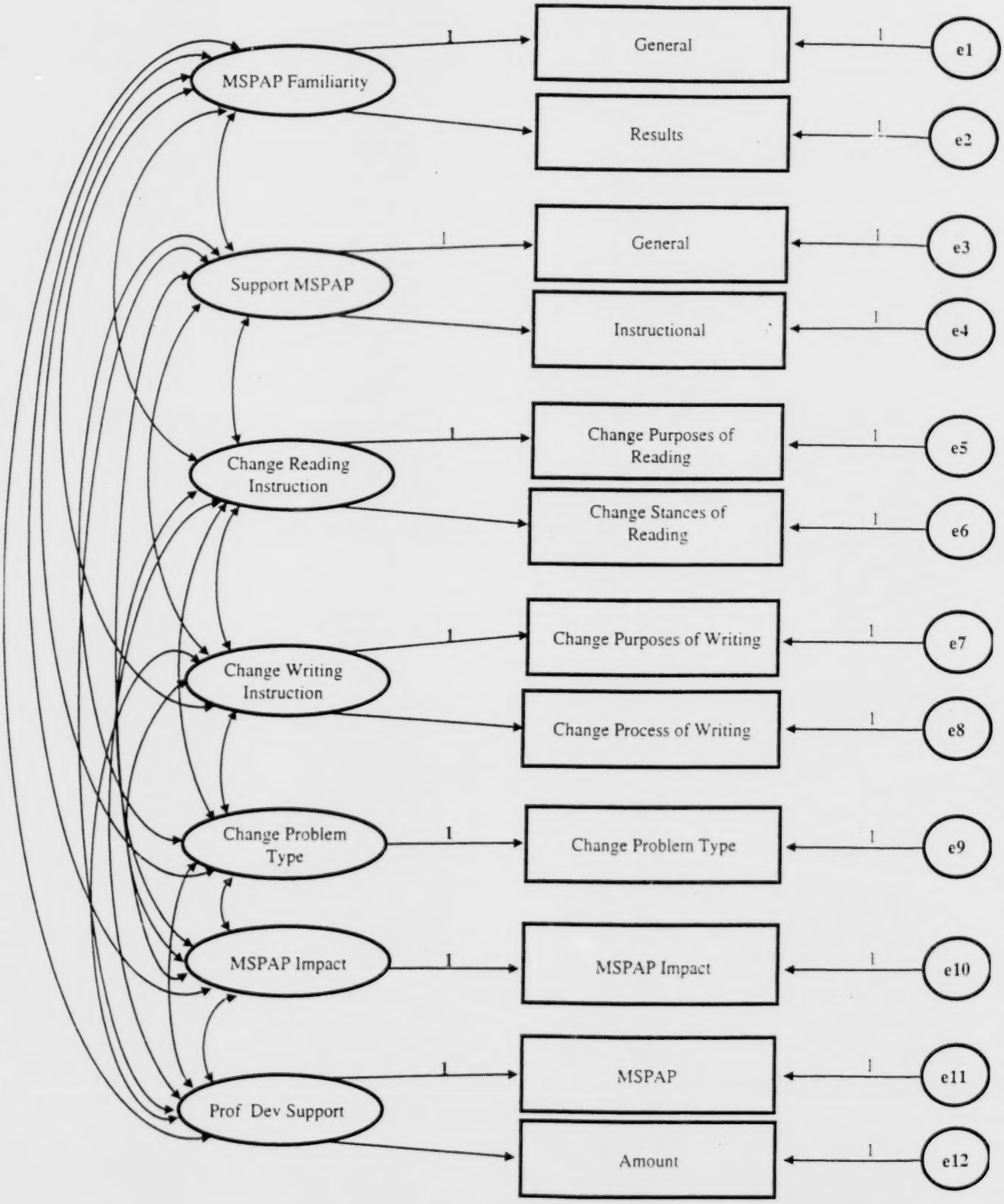
**Figure 4. Four-Factor Model Excluding Current Instruction Measures**






BEST COPY AVAILABLE

**Figure 5. Seven-Factor Model Excluding Current Instruction Measures**






BEST COPY AVAILABLE

Results for the Analyses Excluding the Instructional Change Measures. As indicated in Table 3, for the on-grade analyses excluding the instructional change measures, the one-factor model and the four-factor model did not fit the data as evidenced by the significant chi-square. The seven-factor model fit the data as evidenced by the nonsignificant chi-square statistic, the RMSEA, and the NFI. Three of the covariances among the factors were not significant including the relationships between Support MSPAP with Current Reading Instruction, Support MSPAP with Current Writing Instruction, and Support MSPAP with Current Problem Type .

Table 3
Confirmatory Factor Analysis Excluding Instruction/Assessment Change Measures – Language Arts Teacher Questionnaire

| | $\chi^2$ | df | P | RMSEA | NFI |
|---|---|---|---|---|---|
| On-grade (n=247) | | | | | |
| 1-factor model | 398.555 | 54 | .000 | .161 | .530 |
| 4-factor model | 131.337 | 48 | .000 | .084 | .845 |
| 7-factor model | 45.752 | 35 | .105 | .035 | .946 |
| Off-grade (n=176) | | | | | |
| 1-factor model | 306.917 | 54 | .000 | .164 | .552 |
| 4-factor model | 132.671 | 48 | .000 | .100 | .806 |
| 7-factor model | 44.896 | 35 | .122 | .040 | .934 |
| On and off grade | | | | | |
| 7-factor model | | | | | |
| Constrained | 164.526 | 116 | .002 | .032 | .893 |
| Unconstrained | 90.659 | 70 | .049 | .026 | .941 |

These analyses were also conducted for the off-grade levels (2, 4, 7), combined. Three similar models, excluding the instructional change measures, were estimated to determine whether the underlying structure of the teacher questionnaire was similar for the on- and off- grades. The seven-factor model for the off-grade levels, which excluded the instructional change measures, fit the data as evidenced by the nonsignificant chi-square statistic, the RMSEA, and the NFI presented in Table 3[1]. Four

[1] The underlying structure of the language arts teacher questionnaire was hypothesized to be similar to that of the mathematics teacher questionnaire presented in Lane, Parke and Stone (1998). In particular the items regarding MSPAP familiarity, Support of MSPAP, and Professional Development Support were identical. However, an item that was used in the confirmatory factor analysis for the math data was omitted in the analyses for the language arts data presented in Table 3. The item that was omitted was in the Instructional Support measure. When this item was used in the analyses of the language arts data for the off-grade, the seven-factor model did not fit the data when the instructional change measures were excluded ($\chi^2$ (35, N=176) = 55.328, p =.016; RMSEA = .058; and NFI = .917) and when the item was used in the analysis for the off-grade that excluded the current instruction measures there was a negative error variance associated with the instructional support measure. Thus, this particular item was excluded from the

of the covariances among the factors were not significant including the relationship between Support MSPAP with Professional Development, Support MSPAP with MSPAP Familiarity, Support MSPAP with Current Reading Instruction, and Support MSPAP with Current Problem Type.

A third set of analyses was conducted to determine whether the parameters could be constrained across the on- and off-grades for the seven factor model. The results are provided in Table 3. The difference $\chi^2$ of 73.867 with 46 df was significant (p = .006), indicating that the additional parameters estimated under the unconstrained model improved on model data fit as offered by the constrained model. Thus, the parameters could not be constrained across the two groups. Table 4 provides the unstandardized regression coefficients, their standard errors, and the significance tests for the seven-factor model with the parameters unconstrained for both the on- and off-grades. The 1's in the column for the unstandardized regression coefficients denote the necessary constraints to attain model identification.

Table 4
Regression Coefficients and Significance Tests for Confirmatory Factor Model with Seven Factors Excluding Instruction/Assessment Change Measures (On and Off Grade Parameters Unconstrained) – Language Arts Teacher Questionnaire

| Dimension and Measure | Unstandardized Regression Coefficients | | SE | | t | |
|---|---|---|---|---|---|---|
| | On | Off | On | Off | On | Off |
| MSPAP Familiarity | | | | | | |
| General | 1.000 | 1.000 | | | | |
| Results | 1.732 | 1.222 | .176 | .151 | 9.832* | 8.114* |
| Support MSPAP | | | | | | |
| General | 1.000 | 1.000 | | | | |
| Instruction | 1.274 | 1.502 | .197 | .423 | 6.464* | 3.548* |
| Current Reading Inst/Assess. | | | | | | |
| Reading Purpose | 1.000 | 1.000 | | | | |
| Reading Stance | .986 | 1.337 | .169 | .183 | 5.843* | 7.317* |
| Current Writing Inst/Assess. | | | | | | |
| Writing Purpose | 1.000 | 1.000 | | | | |
| Writing Process | .545 | .649 | .090 | .143 | 6.085* | 4.545* |
| Problem Types | 1.000 | 1.000 | | | | |
| MSPAP Impact | 1.000 | 1.000 | | | | |
| Professional Dev. Support | | | | | | |
| MSPAP | 1.000 | 1.000 | | | | |
| Amount | .888 | .871 | .121 | .141 | 7.349* | 6.192* |

Note: *p < .01

language arts on- and off-grade data sets as reported in Table 3. This is in contrast to the models that fit

Results for the Analyses Excluding the Current Instructional Measures. Similar results were found for the on-grade analyses that excluded the current instructional measures. The one-factor model and the four-factor model did not fit the data as evidenced by the significant chi-square statistics in Table 5. The seven-factor model fit the data as evidenced by the nonsignificant chi-square statistic, the RMSEA, and the NFI. Three covariances among the factors were not significant including the relationships between Professional Development Support with Change in Reading Instruction, Professional Development Support with Change in Writing Instruction, and MSPAP Familiarity with Change in Reading Instruction.

Table 5
Confirmatory Factor Analysis Excluding Current Instruction/Assessment Measures – Language Arts Teacher Questionnaire

| | $\chi^2$ | df | p | RMSEA | NFI |
|---|---|---|---|---|---|
| On-grade (n=185) | | | | | |
| 1-factor model | 369.469 | 54 | .000 | .178 | .548 |
| 4-factor model | 87.200 | 48 | .000 | .067 | .893 |
| 7-factor model | 46.464 | 35 | .093 | .042 | .943 |
| Off-grade (n= 126) | | | | | |
| 1-factor model | 248.078 | 54 | .000 | .170 | .548 |
| 4-factor model | 96.325 | 48 | .000 | .090 | .838 |
| 7-factor model | 34.033 | 35 | .515 | .000 | .941 |
| On and off grade | | | | | |
| 7-factor model | | | | | |
| Constrained | 121.265 | 113 | .281 | .015 | .913 |
| Unconstrained | 80.501 | 70 | .183 | .022 | .942 |

These analyses were also conducted for the off-grade levels (2, 4, 7) combined. Three models, excluding the current instructional measures, were estimated to determine whether the underlying structure of the teacher questionnaire was similar for the on- and off-grades. Similar to the on-grade levels, the one- and four-factor models for the off-grades did not fit the data as evidenced by the significant chi-square statistic in Table 5. The seven factor model for the off-grade levels did fit the data as evidenced by the non-significant chi-square statistic, the RMSEA, and the NFI presented in Table 5[2].

---

the data for the mathematics questionnaire (Lane, Parke, & Stone, 1998).

[2] It should be noted that a model was tested that included both the current instruction measures and the instructional change measures (i.e., all ten dimensions) for both the on- and off- grade data sets. However, the model did not fit the data for both the on- and off-grades ($\chi^2$ (77, N= 185) = 123.817, p =.001, RMSEA = .057, and NFI = .894; and $\chi^2$ (77, N=126) =127.972, p < .001, RMSEA = .073, and NFI = .870, respectively). This is in contrast with the math data reported in Lane, Parke and Stone (1998). The model that included both the current instruction measures and the instructional change measures fit the mathematics data.

Six of the covariances among the factors were not significant including Support MSPAP with each of the following: Professional Development Support, MSPAP Familiarity, Change in Reading Instruction, Change in Writing Instruction, and Change in Problem Type. In addition, the covariance between Change in Problem Type and MSPAP Familiarity was not significant.

Another set of analyses was conducted to determine whether the parameters could be constrained across the on- and off-grades for the seven factor model, including instructional change. The results are provided in Table 5. The difference $\chi^2$ of 40.764 with 43 df was not significant (p=.569) indicating that the additional parameters estimated under the unconstrained model did not improve on model data fit. Thus, the parameters could be constrained across the two groups. Table 6 provides the unstandardized regression coefficients, their standard errors, and the significance tests for the seven-factor model for the combined on- and off-grade levels.

Table 6
Regression Coefficients and Significance Tests for Confirmatory Factor Model with Seven Factors Excluding Current Instruction/Assessment Measures (On and Off Grade Parameters Constrained) – Language Arts Teacher Questionnaire

| Dimension and Measure | Unstandardized Regression Coefficients | SE | t |
|---|---|---|---|
| MSPAP Familiarity | | | |
| General | 1.000 | | |
| Results | 1.520 | .142 | 10.687* |
| Support MSPAP | | | |
| General | 1.000 | | |
| Instruction | 1.380 | .230 | 5.992* |
| Change Reading Inst/Assess. | | | |
| Reading Purpose | 1.000 | | |
| Reading Stance | 1.083 | .083 | 12.992* |
| Change Writing Inst/Assess. | | | |
| Writing Purpose | 1.000 | | |
| Writing Process | .896 | .081 | 11.057* |
| Problem Types | 1.000 | | |
| MSPAP Impact | 1.000 | | |
| Professional Dev. Support | | | |
| MSPAP | 1.000 | | |
| Amount | 1.032 | .117 | 8.830* |

Note: *p < .01

**Analysis of Variance for the Language Arts Questionnaire Data**

Results for the Teacher Questionnaire. An analysis of variance was conducted for each of the ten dimensions with the between-subjects effect being the grade and the dependent measure being the teacher composite mean score on the dimension[3]. Descriptive data for the dimensions are provided in Table 7. The range on the questionnaire item scale is 1 - 4, with the more positive responses being at the upper end of the scale. Overall, the mean scores are at the upper end of the scale.

Table 7
Descriptive Data for the Ten Dimensions – Language Arts Teacher Questionnaire

| Dimension | | Off-Elem (2nd/4th) (n=117) | On-Elem (3rd/5th) (n=156) | Off-Middle (7th) (n=59) | On-Middle (8th) (n=91) |
|---|---|---|---|---|---|
| MSPAP | mean | 3.160 | 3.326 | 2.977 | 3.151 |
| Familiarity | sd | .562 | .561 | .572 | .616 |
| Support | mean | 2.563 | 2.578 | 2.723 | 2.558 |
| MSPAP | sd | .620 | .668 | .513 | .655 |
| Current Reading | mean | 3.503 | 3.454 | 3.266 | 3.504 |
| Inst/Assess. | sd | .409 | .439 | .608 | .472 |
| Current Writing | mean | 3.544 | 3.620 | 3.525 | 3.596 |
| Inst/Assess. | sd | .404 | .398 | .427 | .385 |
| Current Problem | mean | 3.123 | 3.170 | 2.839 | 2.862 |
| Type | sd | .568 | .467 | .527 | .530 |
| Change Reading | mean | 3.091 | 3.064 | 2.730 | 2.934 |
| Inst/Assess.* | sd | .4840 | .474 | .581 | .502 |
| Change Writing | mean | 3.251 | 3.141 | 2.989 | 2.916 |
| Inst/Assess.* | sd | .499 | .506 | .486 | .450 |
| Change Problem | mean | 3.019 | 2.900 | 2.792 | 2.821 |
| Type* | sd | .506 | .495 | .464 | .453 |
| MSPAP | mean | 3.163 | 3.353 | 2.770 | 2.950 |
| Impact | sd | .550 | .554 | .513 | .660 |
| Professional | mean | 2.853 | 2.942 | 2.672 | 2.760 |
| Dev. Support | sd | .598 | .637 | .657 | .639 |

*The sample sizes for change reading, change writing, and change problem type were 88, 111, 38, and 74 for off-elementary, on-elementary, off-middle, and on-middle, respectively.

Table 8 provides a summary of the analysis of variance results. As indicated in the table, there were significant grade differences for eight of the dimensions: MSPAP Familiarity, Current Reading Instruction, Change in Reading Instruction, Change in Writing Instruction, Current Problem Type,

[3] A multivariate analysis of variance was not conducted because the current instruction measures and the instructional change measures could not be included in the same confirmatory factor model.

Change in Problem Type, MSPAP Impact on Instruction, and Professional Development Support. There were no significant differences among grades for Support MSPAP and Current Writing Instruction.

Table 8
ANOVA Results for Ten Dimensions – Language Arts Teacher Questionnaire

| Dimension | df | F | p | $r^2$ |
|---|---|---|---|---|
| MSPAP Familiarity | 3 | 5.854 | .001 | .033 |
| Support MSPAP | 3 | 1.039 | .375 | .000 |
| Current Reading Inst/Assess. | 3 | 3.980 | .008 | .021 |
| Current Writing Inst./Assess. | 3 | 1.254 | .290 | .002 |
| Current Problem Type | 3 | 10.803 | .000 | .065 |
| Change Reading Inst/Assess. | 3 | 5.753 | .001 | .044 |
| Change Writing Inst/Assess. | 3 | 7.211 | .000 | .057 |
| Change Problem Type | 3 | 3.053 | .029 | .019 |
| MSPAP Impact | 3 | 18.801 | .000 | .112 |
| Professional Dev. Support | 3 | 3.272 | .021 | .016 |

For each of the eight dimensions that were significant, Tukey HSD post-hoc analyses were conducted to determine which differences between composite mean scores were significant. Table 9 provides the results of the post-hoc analyses. In general, an examination of the table indicates that composite mean scores for elementary on- and off-grade teachers tended to be significantly greater than composite mean scores for middle on- and off-grade teachers. For example, on- and off- grade elementary teachers, as compared to on- and off-grade middle school teachers, were more likely to indicate that they place a greater emphasis on reform oriented problem types, as shown by the significant mean differences on the variable, Current Problem Type. As another example, on- and off- grade elementary teachers, as compared to on- and off-grade middle school teachers, were more likely to indicate that MSPAP had a greater impact on their language arts classroom instruction and assessment practices, as evidenced by the significant mean differences on the variable, MSPAP Impact. In addition, on- and off-grade elementary teachers, as compared to off-grade middle school teachers, were more likely to indicate that they place more emphasis on the reading learning outcomes, as evidenced by the mean differences for the dimension, Current Reading Instruction. The same result occurred for Change in Reading Instruction. Further, on- and off- grade elementary teachers, as compared to on-grade middle school teachers, indicated that their emphasis on the learning outcomes for Writing is greater than it was a few years ago, as evidenced by the mean differences for the variable, Change in Writing Instruction. Lastly, on-grade elementary teachers, as compared to middle off-grade teachers were more likely to

indicate that they had received more professional development support regarding MSPAP as evidenced by the mean differences for the dimension, Professional Development Support.

Table 9
Tukey HSD Post-Hoc Results – Language Arts Teacher Questionnaire

| Dimension | Contrast | Mean Difference | SE | p |
|---|---|---|---|---|
| MSPAP Familiarity | 3/5 vs 7 | .265 | .096 | .030 |
| Current Reading Instruction/ | 2/4 vs 7 | .237 | .074 | .008 |
| Assessment | 3/5 vs 7 | .188 | .071 | .041 |
| | 8 vs 7 | .238 | .078 | .012 |
| Current Problem Type | 2/4 vs 7 | .285 | .083 | .003 |
| | 2/4 vs 8 | .265 | .073 | .002 |
| | 3/5 vs 7 | .331 | .079 | .000 |
| | 3/5 vs 8 | .308 | .068 | .000 |
| Change Reading Instruction/ | 2/4 vs 7 | .361 | .097 | .001 |
| Assessment | 3/5 vs 7 | .334 | .093 | .002 |
| Change Writing Instruction/ | 2/4 vs 7 | .262 | .095 | .029 |
| Assessment | 2/4 vs 8 | .335 | .077 | .000 |
| | 3/5 vs 8 | .225 | .073 | .011 |
| Change Problem Type | 2/4 vs 8 | .198 | .076 | .047 |
| MSPAP Impact | 2/4 vs 7 | .393 | .091 | .000 |
| | 2/4 vs 8 | .212 | .080 | .040 |
| | 3/5 vs 2/4 | .190 | .070 | .033 |
| | 3/5 vs 7 | .583 | .087 | .000 |
| | 3/5 vs 8 | .402 | .075 | .000 |
| Professional Development | 3/5 vs 7 | .270 | .096 | .026 |

There was only one difference between mean scores for elementary on- and off-grades. Elementary on-grade teachers, as compared to elementary off-grade teachers, were more likely to indicate that MSPAP had a greater impact on classroom instruction, as evidenced by the mean difference for the variable, MSPAP Impact. However, the mean difference is relatively small.

There was also only one difference between mean scores for middle on- and middle off-grades. Middle on-grade teachers, as compared to middle off-grade teachers, were more likely to indicate that they place more emphasis on the Learning Outcomes for Reading, as evidenced by the mean differences for the variable, Current Reading Instruction. Again, however, the mean difference is small.

As indicated in Table 8, the adjusted $r^2$ value is relatively small for each of the significant variables indicating that grade accounts for only a small percentage of the variance. Thus, although the above results indicate significant mean differences among grades, the actual differences are small. The greatest differences are for the MSPAP Impact dimension. In particular, on-grade elementary teachers indicated

to a greater extent that MSPAP had an impact on their classroom instruction and assessment practices as compared to middle on- and off-grades (mean differences of .583 and .402, respectively).

Results for the Principal and Student Language Arts Questionnaire. Elementary and middle school principals were asked to respond to some of the same items as in the teacher questionnaire. Table 10 provides elementary and middle school principal mean scores on four of the dimensions: MSPAP Familiarity, Support MSPAP, MSPAP Impact, and Professional Development Support. This table also provides corresponding mean scores for the teachers. It should be noted that some of the mean scores for the teachers in this table are somewhat different than the mean scores provided in Table 7. This is because the scores in Table 10 are based only on the items that were the same for the principals and the teachers. For the dimensions, MSPAP Familiarity and Support MSPAP, the items were the same for both teachers and principals. For the dimensions, MSPAP Impact and Professional Development Support, the principals had fewer items than the teachers and consequently the teacher mean scores in Table 10 are based on a smaller number of items than those reported in Table 7. As indicated in Figure 1, the principals were administered four of the six items for MSPAP Impact. For Professional Development Support, they were administered only the two items in the measure, Amount of Support, and not the items in the measure, MSPAP Related Professional Development Support.

Table 10
Descriptive Data for Four Dimensions - Teacher and Principal Language Arts Questionnaire

| | | Teacher | | | | Principal | |
|---|---|---|---|---|---|---|---|
| Dimension | | Off-Elem ($2^{nd}/4^{th}$) (n=117) | On-Elem ($3^{rd}/5^{th}$) (n=156) | Off-Middle ($7^{th}$) (n=59) | On-Middle ($8^{th}$) (n=91) | Elem (n=56) | Middle (n=27) |
| MSPAP Familiarity | mean | 3.160 | 3.326 | 2.977 | 3.151 | 3.613 | 3.475 |
| | sd | .562 | .561 | .572 | .616 | .397 | .369 |
| Support MSPAP | mean | 2.563 | 2.578 | 2.723 | 2.558 | 3.128 | 3.203 |
| | sd | .620 | .668 | .513 | .655 | .565 | .482 |
| MSPAP Impact | mean | 3.182 | 3.352 | 2.792 | 2.965 | 3.488 | 3.278 |
| | sd | .649 | .617 | .600 | .737 | .544 | .462 |
| Professional Dev Support | mean | 2.906 | 2.904 | 2.723 | 2.731 | 3.210 | 3.080 |
| | sd | .788 | .797 | .779 | .793 | .551 | .657 |

An analysis of variance was conducted for each of the four dimensions, with the between-subjects effect being elementary vs. middle and the dependent measure being the principal composite mean score on the dimension. Table 11 provides a summary of the results of the analyses. As indicated in the table,

there were no significant differences on any of the dimensions between elementary and middle school principals[4].

Table 11
ANOVA Results for Four Dimensions – Teacher and Principal Language Arts Questionnaires

| | Teacher | | | | Principal | | | |
|---|---|---|---|---|---|---|---|---|
| Dimension | df | F | p | $r^2$ | df | F | p | $r^2$ |
| MSPAP Familiarity | 3 | 5.854 | .001 | .033 | 1 | 2.238 | .139 | .015 |
| Support MSPAP | 3 | 1.039 | .375 | .000 | 1 | .347 | .557 | -.008 |
| MSPAP Impact | 3 | 13.462 | .000 | .081 | 1 | 2.991 | .088 | .024 |
| Professional Dev. Support | 3 | 1.574 | .195 | .004 | 1 | .886 | .349 | -.001 |

In general, the principal composite mean scores were higher than the teacher composite mean scores on the dimensions as indicated in Table 10. For each of the four dimensions, an analysis of variance was conducted, with the between-subjects effect being teacher vs. principal. As indicated in Table 12, all tests were significant. Elementary and middle school principals, as compared to elementary and middle school teachers of Language Arts, indicated that they were more familiar with MSPAP, they were more supportive of MSPAP, MSPAP had a greater impact on classroom instruction and assessment practices, and teachers were receiving more professional development support. It should be noted, however, that the adjusted $r^2$ values are relatively small.

Table 12
ANOVA Results for Four Dimensions – Teacher vs. Principal

| Dimension | df | F | p | $r^2$ |
|---|---|---|---|---|
| MSPAP Familiarity | 1 | 31.122 | .000 | .056 |
| Support MSPAP | 1 | 57.478 | .000 | .101 |
| MSPAP Impact | 1 | 12.238 | .001 | .022 |
| Professional Dev. Support | 1 | 12.574 | .000 | .022 |

[4] Table 11 also presents similar analyses for the teacher data. These results are similar to those presented in Table 8. The results are identical in both tables for MSPAP Familiarity and Support MSPAP because the principals and teachers were given the same items in these dimensions. The teacher results for MSPAP Impact and Professional Development Support are somewhat different because in Table 11 they are based on a fewer number of items. In particular, the test for Professional Development Support was significant for the teachers in Table 8. Recall. this dimension in Table 8 reflects both Amount of Support and MSPAP Related Support, however, the dimension represented in Table 11 reflects only the measure, Amount of Support.

Students in 4th, 5th, 7th, and 8th grade were also asked to respond to some of the same items as in the teacher questionnaire related to the dimensions, Current Reading Instruction, Current Writing Instruction, and Current Problem Type. Classroom-level composite mean scores for each of the grades were obtained on these three dimensions and are provided in Table 13. This table also provides corresponding mean scores for the teachers. It should be noted that mean scores for the teachers in this table are somewhat different than the mean scores provided in Table 7. This is because the scores in Table 13 are based only on the items that were the same for the students and teachers. In particular, for the dimension, Current Reading Instruction, the students were administered only those items that were related to the measure, Purpose of Reading, as indicated in Figure 1. Thus, the items related to the measure, Stances of Reading, are excluded from these analyses. In addition, only teachers with corresponding class (student) data were considered for these analyses.

Table 13

Descriptive Data for Three Dimensions- Teacher and Student Language Arts Questionnaire

| Dimension | | Teacher | | | | Classes (Students) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Off-Elem ($2^{nd}/4^{th}$) (n=48) | On-Elem ($3^{rd}/5^{th}$) (n=55) | Off-Middle ($7^{th}$) (n=47) | On-Middle ($8^{th}$) (n=59) | Off-Elem ($4^{th}$) (n=48) | On-Elem ($5^{th}$) (n=55) | Off- Middle ($7^{th}$) (n=47) | On- Middle ($8^{th}$) (n=59) |
| Current Reading | mean | 3.500 | 3.607 | 3.368 | 3.453 | 3.055 | 3.037 | 2.838 | 2.919 |
| Inst/Assess. | sd | .491 | .450 | .623 | .433 | .284 | .289 | .276 | .289 |
| Current Writing | mean | 3.568 | 3.711 | 3.587 | 3.587 | 2.868 | 2.977 | 2.924 | 3.006 |
| Inst/Assess. | sd | .420 | .341 | .358 | .396 | .202 | .279 | .263 | .233 |
| Current | mean | 3.123 | 3.295 | 2.773 | 2.877 | 2.570 | 2.632 | 2.412 | 2.383 |
| Problem Type | sd | .588 | .415 | .543 | .485 | .327 | .354 | .346 | .283 |

For each of the three dimensions, a one-way analysis of variance, with the between-subjects effect being the grade level was conducted on the class data. As indicated in Table 14, the tests were significant for students[5]. However, the adjusted $r^2$ values are small.

Table 14
ANOVA Results for the Three Dimensions – Teacher and Class (Students)

| | Teacher | | | | Class (Students) | | | |
|---|---|---|---|---|---|---|---|---|
| Dimension | df | F | p | $r^2$ | df | F | p | $r^2$ |
| Current Reading Inst/Assess. | 3 | 2.056 | .107 | .015 | 3 | 6.042 | .000 | .072 |
| Current Writing Inst/Assess. | 3 | 1.597 | .191 | .009 | 3 | 3.158 | .026 | .030 |
| Current Problem Type | 3 | 11.350 | .000 | .130 | 3 | 7.384 | .000 | .084 |

Tukey HSD post-hoc analyses were conducted to determine which differences between mean scores were significant for Current Reading Instruction, Current Writing Instruction, and Current Problem Type. Table 15 provides the results of the post-hoc analyses.

Table 15
Tukey HSD Post-Hoc Results– Teacher and Student Language Arts Questionnaire

| | Teacher | | | | Class (Students) | | | |
|---|---|---|---|---|---|---|---|---|
| Dimension | Contrast | Mean Difference | SE | p | Contrast | Mean Difference | SE | p |
| Current Reading Instruction/Asst* | | | | | 4 vs 7 | .217 | .058 | .001 |
| | | | | | 5 vs 7 | .199 | .057 | .042 |
| Current Writing Instruction/Asst* | | | | | 8 vs 4 | .138 | .048 | .021 |
| Current Problem Type | 2/4 vs 7 | .350 | .104 | .004 | 4 vs 8 | .188 | .064 | .017 |
| | 3/5 vs 7 | .523 | .101 | .000 | 5 vs 7 | .220 | .065 | .004 |
| | 3/5 vs 8 | .419 | .095 | .000 | 5 vs 8 | .250 | .061 | .000 |

Note: * not significant for teachers

Elementary off-grade ($4^{th}$) and on-grade ($5^{th}$), as compared to middle off-grade ($7^{th}$) students, were more likely to indicate that greater emphasis was placed on the learning outcomes for reading (Purposes of Reading) in their language arts classrooms. Middle school on-grade ($8^{th}$) students, as compared to elementary off-grade ($4^{th}$) students were more likely to indicate that greater emphasis was placed on the learning outcomes for writing (Purposes for Writing and Process of Writing) in their language arts

[5] Table 14 also presents similar results for the teacher data. The results for teachers are similar to those presented in Table 8. However, the test for Current Reading Instruction is not significant for the teacher data set reported in Table 14, whereas, it was significant for the teacher data set reported in Table 8. Recall, the dimension reported in

classrooms. Further, elementary off-grade ($4^{th}$) and on-grade ($5^{th}$) students, as compared to middle on-grade ($8^{th}$) students, were more likely to indicate that greater emphasis was placed on reform-oriented problem types in their language arts classroom. The same result was found for elementary on-grade ($5^{th}$) versus middle off-grade ($7^{th}$).

In general, the composite mean scores for classes on the three dimensions were consistently lower than the teacher composite mean scores as indicated in Table 13. For each dimension, a one-way analysis of variance, with the between-subjects effect being the student/teacher, was conducted on the data. The results of these analyses are provided in Table 16. All three tests were significant.

Table 16
ANOVA Results for the Four Dimensions – Teacher vs. Student

| Dimension | df | F | p | $r^2$ |
|---|---|---|---|---|
| Current Reading Inst/Assess. | 1 | 167.890 | .000 | .286 |
| Current Writing Inst/Assess. | 1 | 448.122 | .000 | .517 |
| Current Problem Type | 1 | 138.214 | .000 | .248 |

Teachers, as compared to students, were more likely to indicate that their Language Arts classrooms had a greater emphasis on the learning outcomes for reading (Purposes of Reading) and for writing (Purposes for Writing and Process of Writing). Further, teachers indicate that their Language Arts classroom had a greater emphasis on reform-oriented problem types. The adjusted $r^2$ values are larger than the adjusted $r^2$ values reported in previous analyses. In particular, the adjusted $r^2$ value of .517 for Current Writing Instruction is relatively large, indicating that 51% of the variance in the Current Writing Instruction variable is accounted for by the type of respondent (teacher vs. student).

## Science Teacher Questionnaire

### Confirmatory Factory Analyses for the Science Teacher Questionnaire

The science teacher questionnaire was designed to provide information about six dimensions. They included teachers' familiarity with MSPAP, teachers' support for MSPAP, teachers' instruction and assessment practices in science, change in teacher's instruction and assessment practices in science, MSPAP's impact on instruction and assessment in science, and professional development support for teachers. Subsets of items were grouped according to the following eleven areas (i.e., measures) to reflect the six dimensions:

---

Table 14 reflects only the Purposes of Reading and not the Stances of Reading. Thus, for this smaller data set, teachers across grade levels responded similarly for the measure, Purposes of Reading.

(1) MSPAP Familiarity - General (teachers' general familiarity with MSPAP),

(2) MSPAP Familiarity - Results (teachers' familiarity with MSPAP results),

(3) Support MSPAP – general (teachers' general support for MSPAP),

(4) Support MSPAP - Instruction (teachers' support for MSPAP for instructional purposes),

(5) Current Science Instruction and Assessment - Learning Outcomes (emphasis on science process learning outcomes in instruction and assessment),

(6) Current Science Instruction and Assessment - Problem Type (emphasis on reform-oriented problem types in science instruction and assessment),

(7) Change Science Instruction and Assessment - Learning Outcomes (change in emphasis on science process learning outcomes in instruction and assessment),

(8) Change Science Instruction and Assessment - Problem Type (change in emphasis on reform-oriented problem types in science instruction and assessment),

(9) MSPAP's Impact (MSPAP's impact on instruction and assessment),

(10) Professional Development Support - MSPAP (professional development activities related to MSPAP), and

(11) Professional Development Support - Amount (amount of professional development activities).

The measures for Current Science Instruction and Assessment and Change in Science Instruction and Assessment reflect the Maryland Learning Outcomes (MLO's) in science.

Teacher mean scores were obtained for each of these eleven measures. The majority of the items on the questionnaire had a four-point Likert scale. For those items that had more than a four-point scale, the responses were recoded to a four-point scale. Teacher data were excluded for those cases in which teachers had left blank more than 25% of the items on any one of the eleven subsets of items.

Figure 5 provides the final set of items for each of ten measures and the hypothesized dimension underlying each of the measures. One of the measures, General Support of MSPAP, is not reflected in this figure. This is because a model including this measure would not fit the data for all of the data sets as discussed in the forthcoming presentation of the results. Coefficient alpha reliability estimates for these ten measures for both on- and off-grade data sets ranged from .697 to .938.

Figure 6
Hypothesized Dimensions, Measures, and Science Teacher Questionnaire Items

| Dimension/ Measure | Teacher Questionnaire Item |
|---|---|
| MSPAP Familiarity | |
| General | To what extent are you familiar with each of the following? |
| | purpose of MSPAP* |
| | format of MSPAP tasks* |
| | content and skills assessed by MSPAP* |
| | how to prepare students for MSPAP* |
| Results | how to interpret and use MSPAP results to improve classroom instruction and assessment* |
| | how to explain MSPAP results to students and/or parents* |
| Support MSPAP | |
| Instruction | MSPAP is a useful tool for helping me make positive changes in my instruction.* |
| | MSPAP is a useful tool for making positive changes in instruction for those teachers who are resistant to change.* |
| | Results of MSPAP provide useful information for making inferences about school improvement.* |
| Current Science Instruction and Assessment | |
| Learning Outcomes | About how often have you asked your students to work on classroom activities that emphasize the integration of knowledge for the purpose of understanding the sciences in the context of inquiry, technology, science in personal and social perspectives, and history and nature of science this year (1997-98)? |
| | About how much emphasis have you placed on each of the following learning outcomes (LO) in your science instruction this year (1997-98)? |
| | nature of science** |
| | habits of mind ** |
| | process of science** |
| | applications of science** |
| | About how much emphasis have you placed on each of the following LO in your science assessment this year (1997-98)? |
| | nature of science |
| | habits of mind |
| | process of science |
| | applications of science |
| Problem Type | About how often have you used each of the following types of activities in your science instruction this year (1997-98)? |
| | questions requiring explanations** |
| | written reports (including lab reports)** |
| | experiments (hands-on investigations & analysis of findings)** |
| | making models or other physical representations |
| | activities that take a few days to complete** |



BEST COPY AVAILABLE

Figure 6
Hypothesized Dimensions, Measures, and Science Teacher Questionnaire Items – Continued

| | |
|---|---|
| Problem Type (cont.) | activities that integrate other subjects into science** |
| | activities relating science concepts to students lives** |
| | About how often have you used each of the following types of science assessment questions and activities this year (1997-98)? |
| | questions requiring explanations** |
| | written reports (including lab reports) |
| | experiments (hands-on investigations and analysis of findings) |
| | making models or other physical representations |
| | activities that integrate other subjects into science |
| | activities relating science concepts to students lives |
| | journal entries** |
| | group activities** |
| Change Science Instruction and Assessment | |
| Learning Outcomes | How has the emphasis on classroom activities that emphasize the integration of knowledge for the purpose of understanding the sciences in the context of inquiry, technology, science in personal and social perspectives, and history and nature of science changed from 1992-93 to 1997-98? |
| | How has the emphasis for each learning outcome in your science instruction changed from 1992-93 to 1997-98? |
| | nature of science |
| | habits of mind |
| | process of science |
| | applications of science |
| | How has the emphasis for each learning outcome in your science assessment changed from 1992-93 to 1997-98? |
| | nature of science |
| | habits of mind |
| | process of science |
| | applications of science |
| Problem Type | How has the emphasis for each activity type used in your science instruction changed from 1992-93 to 1997-98? |
| | questions requiring explanations |
| | written reports (including lab reports) |
| | experiments (hands-on investigations and analysis of findings) |
| | making models or other physical representations |
| | activities that take a few days to complete |
| | activities that integrate other subjects into science |
| | activities relating science concepts to students lives |

BEST COPY AVAILABLE

Figure 6
Hypothesized Dimensions, Measures, and Science Teacher Questionnaire Items – Continued

| | |
|---|---|
| Problem Type (cont.) | How has the emphasis for each type of assessment question and activity changed from 1992-93 to 1997-98? |
| | questions requiring explanations |
| | written reports (including lab reports) |
| | experiments (hands-on investigations and analysis of findings) |
| | making models or other physical representations |
| | activities that integrate other subjects into science |
| | activities relating science concepts to students lives |
| | journal entries |
| | group activities |
| MSPAP Impact | |
| | About how much has MSPAP influenced your science classroom activities this year (1997-98)? |
| | To what extent has MSPAP influenced you to make positive changes in your science instruction?* |
| | To what extent has MSPAP influenced you to make positive changes in your science assessment?* |
| | About how often do you ask your students to solve science tasks similar to those on MSPAP? |
| | To what extent have you focused on the following strategies in preparing your students for MSPAP? |
| | increasing the use of MSPAP-like tasks in regular instruction* |
| | increasing the match between the content of instruction and the content of MSPAP* |
| Professional Development Support | |
| Focus on MSPAP | To what extent did staff development activities address the following? |
| | purpose of MSPAP |
| | format of MSPAP tasks |
| | content and skills assessed by MSPAP |
| | how to prepare students for MSPAP |
| | how to interpret and use MSPAP results to improve instruction and assessment |
| | how to explain MSPAP results to students/parents |
| Amount of Support | To what extent have you had the necessary support and/or resources to enable you to make instructional and assessment changes to better reflect what is expected of students in MSPAP and the Maryland Learning Outcomes? |
| | support for instruction* |
| | support for assessment* |

Note: The * items indicate that principals were administered similar items, but from the principal's perspective. The ** items indicate that students were administered similar items, but from the student's perspective.

Two sets of analyses were conducted. For each set of analyses, a maximum likelihood estimation procedure was used to estimate three hierarchical models using AMOS (Arbuckle, 1997). The first set excluded teacher mean scores related to change in their instruction and assessment, Change Science Instruction-Learning Outcomes and Change Science Instruction-Problem Type. The second set of analyses excluded teacher mean scores related to their current instruction and assessment, Current Science Instruction-Learning Outcomes and Current Science Instruction-Problem Type. Teachers answered the questions with respect to instructional change only if they were teachers in Maryland since the 1993-94 school year. Thus, the second set of analyses is based on a smaller sample size than the first set of analyses.

For the analyses excluding the instructional change measures, the first model that was estimated provided a test for the hypothesis that one factor accounted for the interrelations among the teacher mean scores for eight measures. The second model that was estimated provided a test for the hypothesis that four factors accounted for the interrelationships. The third model that was estimated, the hypothesized model, provided a test for the hypothesis that five factors accounted for the interrelationships as specified in Figure 7. For the analyses excluding the current instruction measures, similar models were estimated and the five-factor model is shown in Figure 8. In these analyses the instructional change measures replaced the current instruction measure.

The analyses were done for the on-grade levels (3, 5, 8) combined and the off-grade levels (2, 4, 7) combined. The sample sizes for the analyses excluding the instructional change measures were 296 for the on-grade and 261 for the off-grade. The sample sizes for the analyses excluding the current instruction measures were 177 for the on-grade and 160 for the off-grade.

**Figure 7 Five-Factor Model Excluding Change Instruction Measures**

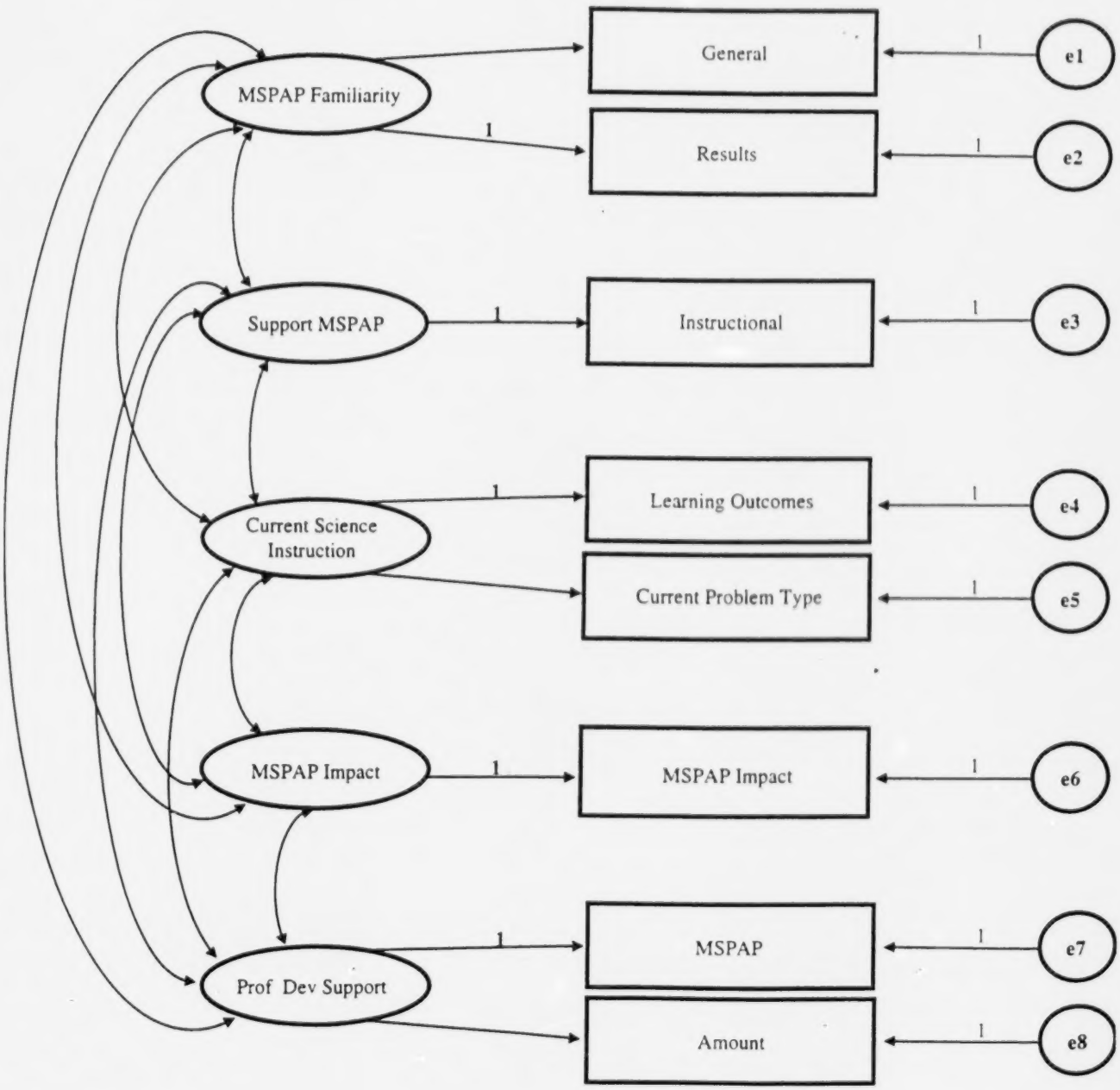
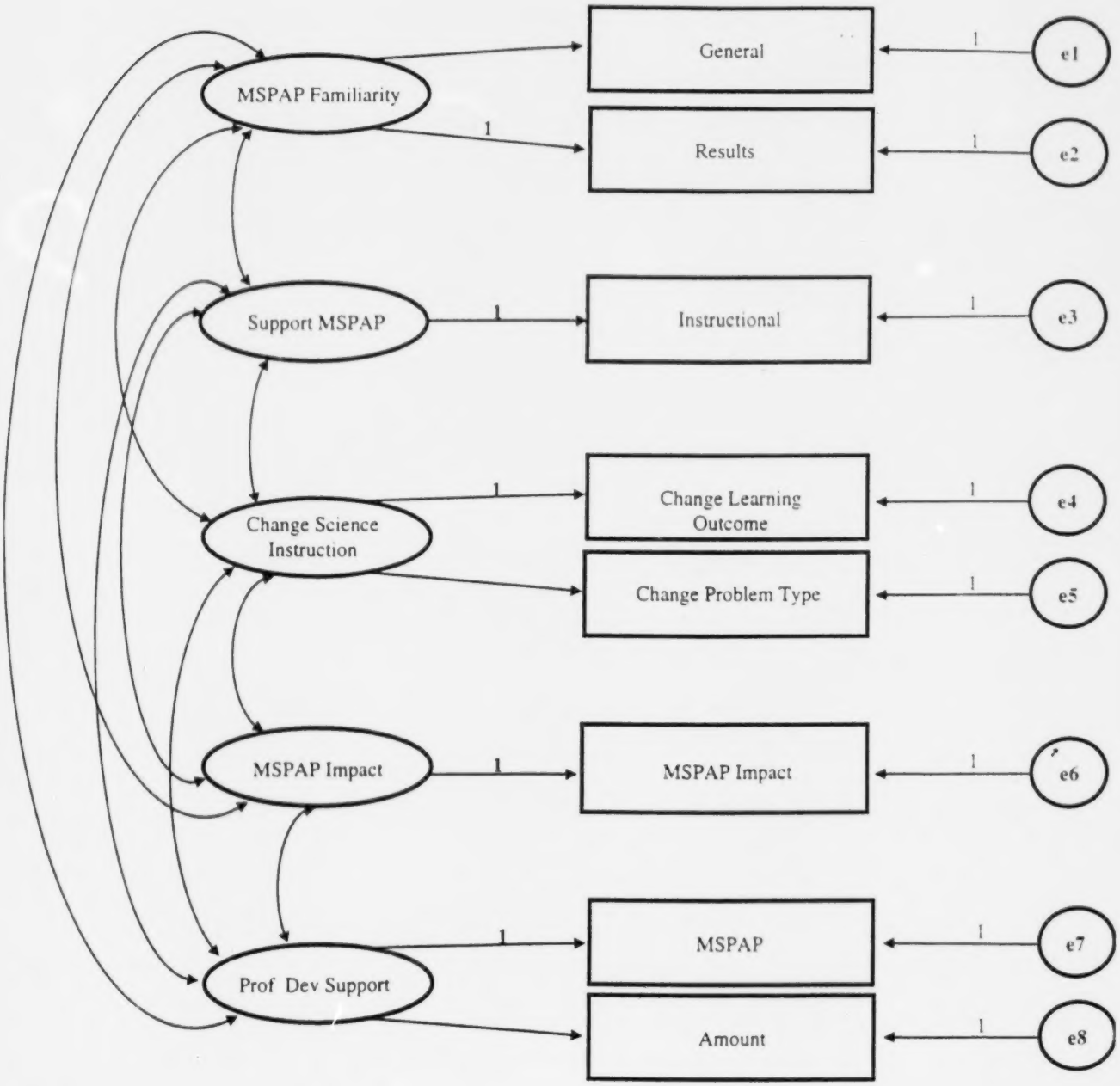


**Figure 8 . Five-Factor Model Excluding Current Instruction Measures**




BEST COPY AVAILABLE

Results for the Analyses Excluding the Instructional Change Measures. As indicated in Table17, for the on-grade analyses excluding the instructional change measures, the one-factor model and the four-factor model did not fit the data as evidenced by the significant chi-square. The five-factor model fit the data as evidenced by the nonsignificant chi-square statistic, the RMSEA, and the NFI. All covariances among the factors were significant.

Table 17
Confirmatory Factor Analysis Excluding Instruction/Assessment Change Measures – Science Teacher Questionnaire

| | $\chi^2$ | df | P | RMSEA | NFI |
|---|---|---|---|---|---|
| On-grade (n=296) | | | | | |
| 1-factor model | 160.142 | 20 | .000 | .154 | .821 |
| 4-factor model | 73.475 | 15 | .000 | .115 | .918 |
| 5-factor model | 10.441 | 12 | .577 | .000 | .988 |
| Off-grade (n=261) | | | | | |
| 1-factor model | 172.129 | 20 | .000 | .171 | .796 |
| 4-factor model | 64.189 | 15 | .000 | .112 | .924 |
| 5-factor model | 13.512 | 12 | .333 | .022 | .984 |
| On and off grade | | | | | |
| 5-factor model | | | | | |
| Constrained | 54.338 | 48 | .246 | .015 | .969 |
| Unconstrained | 23.954 | 24 | .464 | .000 | .986 |

These analyses were also conducted for the off-grade levels (2, 4, 7), combined. Three similar models, excluding the instructional change measures, were estimated to determine whether the underlying structure of the teacher questionnaire was similar for the on- and off- grades. The five-factor model for the off-grade levels, which excluded the instructional change measures fit the data as evidenced by the nonsignificant chi-square statistic, the RMSEA, and the NFI in Table17[6]. All covariances among the factors were significant.

[6] The underlying structure of the science teacher questionnaire was hypothesized to be similar to that of the mathematics teacher questionnaire presented in Lane, Parke, and Stone (1998). In particular the items regarding MSPAP familiarity, Support of MSPAP, and Professional Development Support were identical. A model was tested that included both of the Support MSPAP measures for both on- and off- grade data sets. However, based on the $\chi^2$, the model did not fit the data for the off-grade ($\chi^2$ (18, N=261)=32.086, p=.001; RMSEA=.055; and NFI=.967). The model did fit the data for the on-grade ($\chi^2$ (18, N=294)=13.228, p=.778; RMSEA=.000; and NFI=.987). Because other analyses were to be conducted on these data, a similar structure was maintained for the on- and off-grades. The general support measure was eliminated from both data sets. This is in contrast to the models that fit the math data reported in Lane, Parke, and Stone (1998). Both the general and instruction support measures were included.

A third set of analyses was conducted to determine whether the parameters could be constrained across the on- and off-grades for the five-factor model. The results are provided in Table 17. The difference chi-square of 30.384 with 24 df was not significant (p = .172), indicating that the additional parameters estimated under the unconstrained model did not improve on model data fit as offered by the constrained model. Thus, the parameters could be constrained across the two groups. Table 18 provides the unstandardized regression coefficients, their standard errors, and the significance tests for the five-factor model with the parameters constrained across the on- and off-grades. The 1's in the column for the unstandardized regression coefficients denote the necessary constraints to attain model identification.

Table 18
Regression Coefficients and Significance Tests for Confirmatory Factor Model with Five Factors Excluding Instruction/Assessment Change Measures (On and Off Grade Parameters Constrained) – Science Teacher Questionnaire

| Dimension and Measure | Unstandardized Regression Coefficients | SE | t |
|---|---|---|---|
| MSPAP Familiarity | | | |
| General | .711 | .042 | 16.845* |
| Results | 1.000 | | |
| Support MSPAP | | | |
| Instruction | 1.000 | | |
| Current Science Instruction/ Assessment | | | |
| Learning outcomes | 1.000 | | |
| Problem types | .919 | .060 | 15.249* |
| MSPAP Impact | 1.000 | | |
| Professional Dev. Support | | | |
| MSPAP | 1.000 | | |
| Amount | 1.103 | .083 | 13.331* |

Note: *p < .01

Results for the Analyses Excluding Current Instruction Measures

Similar results were obtained for the on-grade analyses that excluded the current instruction measures. The one-factor model and the four-factor model did not fit the data as evidenced by the significant chi-square statistics in Table 19. The five-factor model fit the data as evidenced by the nonsignificant chi-square statistic, the RMSEA, and the NFI. All covariances among the factors were significant.

Table 19
Confirmatory Factor Analysis Excluding Current Instruction/Assessment Measures – Science Teacher Questionnaire

| | $\chi^2$ | df | p | RMSEA | NFI |
|---|---|---|---|---|---|
| On-grade (n=177) | | | | | |
| 1-factor model | 126.429 | 20 | .000 | .174 | .769 |
| 4-factor model | 64.629 | 15 | .000 | .137 | .882 |
| 5-factor model | 7.641 | 12 | .813 | .000 | .986 |
| Off-grade (n=160) | | | | | |
| 1-factor model | 207.018 | 20 | .000 | .243 | .656 |
| 4-factor model | 64.405 | 15 | .000 | .144 | .893 |
| 5-factor model | 12.621 | 12 | .397 | .018 | .979 |
| On and off grade | | | | | |
| 5-factor model | | | | | |
| Constrained | 55.484 | 48 | .213 | .022 | .952 |
| Unconstrained | 20.264 | 24 | .682 | .000 | .982 |

These analyses were also conducted for the off-grade levels (2, 4, 7) combined. Three models, excluding the current instruction measures, were estimated to determine whether the underlying structure of the teacher questionnaire was similar for the on- and off-grades. Similar to the on-grade levels, the one- and four-factor models for the off-grades did not fit the data as evidenced by the significant chi-square statistic in Table 19[7]. The five-factor model for the off-grade levels did fit the data as evidenced by the nonsignificant chi-square statistic, the RMSEA, and the NFI in Table 19. All covariances among the factors were significant.

Another set of analyses was conducted to determine whether the parameters could be constrained across the on- and off-grades for the five-factor model, including instructional change. The results are provided in Table 19. The difference chi-square of 35.220 with 24 df was not significant (p=.065), indicating that the additional parameters estimated under the unconstrained model did not improve on model data fit. Thus, the parameters could be constrained across the two groups. Table 20 provides the unstandardized regression coefficients, their standard errors, and the significance tests for the five-factor model with the parameters constrained across the on- and off-grades. The 1's in the column for the

[7] It should be noted that a model was tested that included both the current instruction measures and the instructional change measures (i.e., all six dimensions) for both on- and off- grade data sets. However, the model did not fit the data for the off-grade ($\chi^2$ (22, N=155)=54.999, p=.000; RMSEA=.099; and NFI=.932). The model did fit the data for the on-grade data set ($\chi^2$ (22, N=176)=27.854, p=.181; RMSEA=.039; and NFI=.960). However, it was decided to maintain the structure from on- to off- grade, so a model excluding the instructional change measures and a model excluding the current instruction measures were estimated for both the on- and off-grades. This is in contrast to the mathematics results reported in Lane, Parke, and Stone (1998). The model that included both the current instruction measures and the instructional change measures fit the mathematics data.

unstandardized regression coefficients denote the necessary constraints to attain model identification. In general, these results suggest that the underlying structure of the teacher questionnaire for the off-grade levels is similar to the structure for the on-grade.

Table 20
Regression Coefficients and Significance Tests for Confirmatory Factor Model with Five Factors Excluding Current Instruction/Assessment Measures (On and Off Grade Parameters Constrained) – Science Teacher Questionnaire

| Dimension and Measure | Unstandardized Regression Coefficients | SE | t |
|---|---|---|---|
| MSPAP Familiarity | | | |
| General | .656 | .053 | 12.485* |
| Results | 1.000 | | |
| Support MSPAP | | | |
| Instruction | 1.000 | | |
| Change Science Instruction/ Assessment | | | |
| Learning outcomes | 1.000 | | |
| Problem types | .859 | .062 | 13.862* |
| MSPAP Impact | 1.000 | | |
| Professional Dev. Support | | | |
| MSPAP | 1.000 | | |
| Amount | 1.057 | .102 | 10.362* |

Note: *p < .01

**Analysis of Variances for the Questionnaire Data**

Results for the Teacher Questionnaire. An analysis of variance was conducted for each of the six dimensions with the between-subjects effect being the grade and the dependent measure being the teacher composite mean score on the dimension[5]. Descriptive data for the dimensions are provided in Table 21. The range on the questionnaire item scale is 1 - 4, with the more positive responses being at the upper end of the scale. Overall, the mean scores are at the upper end of the scale.

Table 22 provides a summary of the analysis of variance results. As indicated in the table, there were significant grade differences for MSPAP Familiarity, Change in Science Instruction, and MSPAP Impact. There were no significant differences among grades for Support MSPAP, Current Science Instruction, and Professional Development Support.

[5] A multivariate analysis of variance was not conducted because the current instruction measures and the instructional change measures could not be included in the confirmatory factor models.

Table 21
Descriptive Data for the Six Dimensions– Science Teacher Questionnaire

| Dimension | | Off-Elem (2nd/4th) (n=184) | On-Elem (3rd/5th) (n=194) | Off-Middle (7th) (n=77) | On-Middle (8th) (n=102) |
|---|---|---|---|---|---|
| MSPAP | mean | 3.162 | 3.356 | 3.039 | 3.262 |
| Familiarity | sd | .630 | .534 | .668 | .627 |
| Support | mean | 2.748 | 2.768 | 2.641 | 2.614 |
| MSPAP | sd | .675 | .608 | .666 | .676 |
| Current Science | mean | 2.988 | 3.033 | 2.998 | 3.110 |
| Instruction/ Asst. | sd | .520 | .463 | .432 | .445 |
| Change Science | mean | 3.101 | 3.101 | 2.851 | 2.945 |
| Instruction/Asst.* | sd | .466 | .397 | .393 | .369 |
| MSPAP | mean | 3.052 | 3.176 | 2.897 | 3.001 |
| Impact | sd | .632 | .623 | .739 | .729 |
| Professional | mean | 2.892 | 2.903 | 2.706 | 2.751 |
| Dev Support | sd | .649 | .630 | .750 | .752 |

Note: *The sample size for change in science instruction/assessment are 106, 111, 54, and 66 for off-elementary, on-elementary, off-middle, and on-middle, respectively.

Table 22
Univariate ANOVA's for the Six Dimensions– Science Teacher Questionnaire

| Dimension | df | F | p | $r^2$ |
|---|---|---|---|---|
| MSPAP Familiarity | 3 | 6.267 | .000 | .028 |
| Support MSPAP | 3 | 1.734 | .159 | .004 |
| Current Science Instruction/Asst. | 3 | 1.544 | .202 | .003 |
| Change Science Instruction/Asst. | 3 | 6.423 | .000 | .046 |
| MSPAP Impact | 3 | 3.790 | .010 | .015 |
| Professional Dev. Support | 3 | 2.518 | .057 | .008 |

For each of the three dimensions that were significant, Tukey HSD post-hoc analyses were conducted to determine which differences between composite mean scores were significant. Table 23 provides the results of the post-hoc analyses. In general, an examination of the table indicates that composite mean scores for elementary on-grade teachers were significantly greater than composite mean scores for middle off-grade teachers for the three dimensions. For example, elementary on-grade teachers, as compared to middle off-grade teachers, were more likely to indicate that their emphasis on the science learning outcomes and on reform-oriented problem types is greater than it was a few years ago, as evidenced by the composite mean difference for the dimension, Change Science Instruction. Elementary on-grade teachers, as compared to middle off-grade teachers, were more likely to indicate that MSPAP had a greater impact on their science instruction and that they were more familiar with

MSPAP as evidenced by the mean differences for the dimensions, MSPAP Impact and MSPAP Familiarity, respectively.

Table 23
Tukey HSD Post-Hoc Analyses – Science Teacher Questionnaire

| Dimension | Contrast | Mean Difference | SE | p |
|---|---|---|---|---|
| MSPAP Familiarity | 3/5 vs 2/4 | .194 | .062 | .010 |
| | 3/5 vs 7 | .317 | .081 | .001 |
| Change Science Instruction/ | 2/4 vs 7 | .251 | .069 | .002 |
| Assessment | 3/5 vs 7 | .250 | .069 | .002 |
| MSPAP Impact | 3/5 vs 7 | .279 | .089 | .010 |

There was only one difference between mean scores for elementary on- and off-grades. Elementary on-grade teachers, as compared to elementary-off grade teachers, were more likely to indicate that they were more familiar with MSPAP. There were no differences between mean scores for middle on- and off-grades. As indicated in Table 22, however, the adjusted $r^2$ value is relatively small for each of the significant variables indicating that grade accounts for only a small percentage of the variance.

Results for the Principal and Student Questionnaire. Elementary and middle school principals were asked to respond to some of the same items as in the teacher questionnaire. Table 24 provides elementary and middle school principal mean scores on four of the dimensions discussed above: MSPAP Familiarity, Support MSPAP, MSPAP Impact, and Professional Development Support.

Table 24
Descriptive Data for Four Dimensions – Teacher and Principal Science Questionnaires

| | | Teacher | | | | Principal | |
|---|---|---|---|---|---|---|---|
| Dimension | | Off-Elem (2nd/4th) (n=184) | On-Elem (3rd/5th) (n=194) | Off-Middle (7th) (n=77) | On-Middle (8th) (n=102) | Elem (n=91) | Middle (n=48) |
| MSPAP | mean | 3.162 | 3.356 | 3.039 | 3.262 | 3.719 | 3.596 |
| Familiarity | sd | .630 | .534 | .668 | .627 | .350 | .453 |
| Support | mean | 2.748 | 2.768 | 2.641 | 2.614 | 3.070 | 3.118 |
| MSPAP | sd | .675 | .608 | .666 | .676 | .645 | .548 |
| MSPAP | mean | 3.163 | 3.282 | 2.965 | 3.019 | 3.417 | 3.240 |
| Impact | sd | .661 | .657 | .816 | .822 | .450 | .622 |
| Professional | mean | 2.905 | 2.974 | 2.727 | 2.843 | 3.275 | 3.260 |
| Dev Support | sd | .790 | .804 | .985 | .841 | .654 | .684 |

This table also provides corresponding mean scores for the teachers. It should be noted that the mean scores for the teachers in this table are somewhat different than the mean scores provided in Table 21. This is because the scores in Table 24 are based only on the items that were the same for the principals and the teachers. For the dimensions, MSPAP Familiarity and Support MSPAP, the items were the same for both teachers and principals. For the dimensions, MSPAP Impact and Professional Development Support, the principals had fewer items than the teachers and consequently the teacher mean scores in Table 24 are based on a smaller number of items than those reported in Table 21. As indicated in Figure 6, the principals were administered four of the six items for MSPAP Impact. For Professional Development Support, they were administered only the two items in the measure, Amount Support, and none of the items in the measure, MSPAP Related Professional Development Support.

An analysis of variance was conducted for each of the four dimensions, with the between-subjects effect being elementary vs. middle and the dependent measure being the principal composite mean score on the dimension. Table 25 provides a summary of the results of the analyses. As indicated in the table, there were no significant differences on any of the dimensions between elementary and middle school principals[9].

Table 25
ANOVA Results for Four Dimensions– Teacher and Principal Science Questionnaires

| | Teacher | | | | Principal | | | |
|---|---|---|---|---|---|---|---|---|
| Dimension | df | F | p | $r^2$ | df | F | p | $r^2$ |
| MSPAP Familiarity | 3 | 6.267 | .000 | .028 | 1 | 3.168 | .077 | .015 |
| Support MSPAP | 3 | 1.734 | .159 | .004 | 1 | .196 | .659 | .006 |
| MSPAP Impact | 3 | 5.097 | .002 | .022 | 1 | 3.709 | .056 | .019 |
| Professional Dev. Support | 3 | 1.763 | .153 | .004 | 1 | .015 | .904 | .007 |

In general, the principal composite mean scores were higher than the teacher composite mean scores on the dimensions as indicated in Table 24. For each of the four dimensions, an analysis of variance was conducted, with the between effect being teacher vs. principal. As indicated in Table 26, all tests were significant. Elementary and middle school principals, as compared to elementary and middle school teachers of science, indicated that they were more familiar with MSPAP, they were more supportive of

[9] Table 25 also presents similar results for the teacher data. These results are similar to those presented in Table 22. The results are identical in both tables for MSPAP Familiarity and Support MSPAP because the principals and teachers were given the same items in these dimensions. The teacher results for MSPAP Impact and Professional Development Support are somewhat different because in Table 25 they are based on a fewer number of items.

MSPAP, MSPAP had a greater impact on classroom instruction and assessment practices, and teachers were receiving more professional development. The adjusted $r^2$ values, however, are small.

Table 26
ANOVA Results for Four Dimensions– Teacher vs. Principal

| Dimension | df | F | p | $r^2$ |
|---|---|---|---|---|
| MSPAP Familiarity | 1 | 66.966 | .000 | .087 |
| Support MSPAP | 1 | 36.778 | .000 | .049 |
| MSPAP Impact | 1 | 9.921 | .002 | .013 |
| Professional Dev. Support | 1 | 24.419 | .000 | .033 |

Students in 4th, 5th, 7th, and 8th grade were also asked to respond to some of the same items as in the teacher questionnaire related to the dimension Current Science Instruction. Classroom-level composite mean scores for each of the grades were obtained on this dimension and are provided in Table 27. This table also provides corresponding mean scores for the teachers. It should be noted that mean scores for the teachers in this table are somewhat different than the mean scores provided in Table 21. This is because in Table 27 the scores are based only on the items that were the same for the students and teachers. As indicated in Figure 6, the students were administered only 13 of the 24 items in the measures, Learning Outcomes and Problem Type.

Table 27
Descriptive Data for the Current Science Instruction Dimension - Teacher and Student Science Questionnaires

| | | Teacher | | | | Classes (Students) | | | |
|---|---|---|---|---|---|---|---|---|---|
| Dimension | | Off-Elem (4th) (n=58) | On-Elem (5th) (n=64) | Off-Mid (7th) (n=65) | On-Mid (8th) (n=77) | Off-Elem (4th) (N= 58) | On-Elem (5th) (N=64) | Off- Mid (7th) (N=65) | On- Mid (8th) (N=77) |
| Current Science Inst/Asst | mean | 3.047 | 3.191 | 3.017 | 3.171 | 2.408 | 2.460 | 2.409 | 2.613 |
| | sd | .573 | .482 | .389 | .380 | .236 | .241 | .274 | .317 |

A one-way analysis of variance, with the between-subjects effect being grade level was conducted on the class data. As indicated in Table 28, the result was significant for students[10]. However, the adjusted $r^2$ is small.

Table 28
ANOVA Results for the Current Science Instruction Dimension – Teacher and Student Science Questionnaires

| | Teacher | | | | Class (Students) | | | |
|---|---|---|---|---|---|---|---|---|
| Dimension | df | F | p | $r^2$ | df | F | p | $r^2$ |
| Current Science Instruction/Asst | 3 | 2.405 | .068 | .016 | 3 | 9.055 | .000 | .084 |

Tukey HSD post-hoc analyses were conducted to determine which differences between mean scores were significant. Table 29 provides the results of the post-hoc analyses. As indicated in the table, middle on-grade students ($8^{th}$) were more likely to indicate that a greater emphasis was placed on the science learning outcomes and reform-oriented problems than middle off-grade students ($7^{th}$), and elementary on- ($5^{th}$) and off-grade students ($4^{th}$).

Table 29
Tukey HSD Post-Hoc Analyses –Student Science Questionnaire

| Dimension | Contrast | Mean Diff. | SE | p |
|---|---|---|---|---|
| Current Science Instruction/Assessment | 8 vs 4 | .205 | .047 | .000 |
| | 8 vs 5 | .153 | .046 | .005 |
| | 8 vs 7 | .204 | .046 | .000 |

Note: There were no significant differences among grades for teachers.

In general, the composite mean scores for students on this dimension were consistently lower than the teacher composite mean scores. A one-way analysis of variance, with the between-subjects effect being the teacher/student was conducted on the data. The univariate test was significant, $F(1) = 359.622$, $p < .001$. This suggests that teachers, as compared to students, were more likely to indicate that their science classroom instruction had a greater emphasis on the learning outcomes. The adjusted $r^2$ value of .405 is larger than the values reported in previous analyses. This result indicates that

[10] Table 27 presents similar results for the teacher data. The results for teachers are similar to those presented in Table 22. Differences in the results are due to a smaller number of items being considered in the analysis presented in Table 27 as well as a smaller sample size for teachers.

approximately 40% of the variance in the Current Science Instruction variable is accounted for by the type of respondent (teacher vs. student).

**Discussion**

This study examined the underlying structure of the teacher language arts and science questionnaires. The dimensions of the questionnaires are related to MSPAP Familiarity, Support MSPAP, Current Instruction and Assessment Practices, Change in Instruction and Assessment Practices, MSPAP Impact and Professional Development Support. It also examined the extent to which the on- and off-grade teachers differed on the dimensions. The extent to which the principals and students differed from the teachers with respect to a subset of the dimensions was also examined.

In general, teacher mean scores for both the Language Arts and Science questionnaires were above the midpoint of the score scale. However, there was variability across the dimensions. As an example, the mean scores for MSPAP Support tended to be lower than the scores for the other dimensions. This is because of the variability in the teacher responses across the items within this dimension. For example, the majority of the language arts teachers indicated that they support MSPAP (52% somewhat support MSPAP and 14% strongly support MSPAP). Further, the majority of the teachers indicated that MSPAP is a useful tool for making positive changes in instruction (55% somewhat and 18% strongly supported MSPAP for this purpose). The majority of the teachers, however, did not support the identification of schools as eligible for reconstitution based on MSPAP and other report card results (68% somewhat or strongly in opposition). In addition, the majority of the teachers did not support the identification of schools for recognition or monetary rewards based on MSPAP and other report card results (71% somewhat or strongly in opposition). Lastly, when asked if the positive impacts of MSPAP outweigh the negative impacts 59% agreed (48% somewhat and 11% strongly agreed) and 37% disagreed (24% somewhat and 13% strongly disagreed). The teacher results were similar for the science questionnaire.

There were several dimensions for which teachers tended to respond relatively high: MSPAP Familiarity, Change in Instruction and Assessment, and MSPAP Impact. For example, when science teachers were asked to indicate the extent to which the emphasis on the science learning outcomes changed in their classrooms since 1992, the majority of the teachers indicated that the emphasis either increased somewhat or greatly (72%, 74%, 71%, and 74% for the nature of science, habits of mind, processes of science, and applications of science, respectively). Most of the remaining teachers indicated that the emphasis stayed about the same. Further, when asked directly to what extent did MSPAP influence them to make positive changes in their science instruction, 76% of the teachers indicated that

MSPAP had a moderate or great amount of influence on their instruction. This may have been facilitated by the amount of professional development support that teachers received. For example, when science teachers were asked to indicate the extent to which they had the necessary support and/or resources to make changes in their instruction to better reflect the Maryland Learning Outcomes and MSPAP, 38% indicated a moderate amount and 32% indicated a great amount. The results were similar for the language arts teacher questionnaire.

Although the teachers indicated that they were making changes consistent with the Maryland Learning Outcomes and MSPAP, 60% of the science teachers indicated that MSPAP had a somewhat negative impact on teacher morale. Whereas, only 18% indicated that MSPAP had a somewhat positive impact on teacher morale. This may be related to the amount of time that is taken away from instruction because of time spent in meetings about MSPAP and time spent preparing for MSPAP (see Lane, Parke, Stone, 1999).

Overall, elementary on- ($3^{rd}/5^{th}$) and off- ($2^{nd}/4^{th}$) grade teachers, as compared to either or both middle on- ($8^{th}$) and off- ($7^{th}$) grade teachers, tended to respond more highly on several of the language arts questionnaire dimensions. The dimensions that had the largest number of differences were related to classroom practices: Current Reading Instruction, Current Problem Type, Change in Reading Instruction, Change in Writing Instruction, MSPAP Impact. For the science questionnaire, fewer differences occurred. Elementary on-grade teachers, as compared to middle off-grade teachers, tended to respond more highly on the dimensions, Change in Science Instruction and MSPAP Impact. For both language arts and science areas, there were minimal differences between on- and off-grade teachers within school type (elementary or middle). In general, when differences occurred they were relatively small (mean differences ranged from .188 to .583 for the language arts questionnaire and .194 to .317 for the science questionnaire).

For both the language arts and science questionnaires, elementary and middle school principals did not differ significantly on the dimensions, MSPAP Familiarity, Support MSPAP, MSPAP Impact, and Professional Development Support. However, the principals mean scores were significantly greater than the teacher mean scores for these dimensions. Principals, as compared to teachers, indicated that they were more familiar with MSPAP and more supportive of MSPAP, that MSPAP was having a greater impact on classroom instruction, and that teachers were receiving more MSPAP related professional development support.

In contrast, teacher mean scores were significantly greater than student mean scores for the three Current Instruction Dimensions on the Language Arts Questionnaire and the Current Science Instruction Dimension on the Science Questionnaire. Teachers indicated that their classroom activities reflect the

learning outcomes and reform-oriented problem types more so than their students. For the student questionnaire, several differences did exist among the grade levels. Students in the elementary on- and off grades, as compared to either or both middle on- and off- grades, indicated that their current reading instruction reflected the learning outcomes and reform oriented problem types to a greater extent. Whereas, middle school 8th graders indicated that their science instruction reflected the learning outcomes to a greater extent than elementary on- and off- grade students and middle off-grade students.

In summary, the evidence in this paper suggests that MSPAP is having an impact on classroom instruction and assessment to some extent. The majority of the language arts and science teachers indicated that they have made changes in their classroom activities to better reflect the Maryland Learning Outcomes and MSPAP. According to the teachers, the impact has been similar between on- and off- grade levels. When differences occurred they were primarily between elementary and middle school teachers: Elementary teachers tended to respond more favorably than middle school teachers. The majority of the teachers also indicated that MSPAP is a useful tool for improving instruction, however, the majority of the teachers indicated that they oppose using MSPAP for identifying schools for rewards or recognition.

**References**


Arbuckle, J.L. (1997). AMOS User's Guide Version 3.6. Chicago: SmallWaters Corporation.

Chudowsky, N. & Behuniak, P. (1997). Establishing the consequential validity for large-scale performance assessments. Paper presented at the annual meeting of the National Council of Measurement, Chicago.

Cronbach, L.J. (1988). Five perspectives on validity argument. In H. Wainer (Ed.), Test validity (pp. 3-17). Hillsdale, NJ: Erlbaum.

Cronbach, L.J. (1989). Construct validation after thirty ears. In R.E. Linn (Ed.), Intelligence: Measurement, theory and public policy (pp. 147-171). Urbana: University of Illinois Press.

Frederiksen, J.R., & Collins, A. (1989). A districts approach to educational testing. Educational Researcher, 18(9), 27-42.

Joreskog, K.G., & Sorbom, D. (1994). LISREL 8 Users Reference Guide. Chicago: Scientific Software.

Kane, M.B, Khattri, N., Reeve, A.L., & Adamson, R.J. (1997). Assessment of Student Performance. Washington, D.C.: Studies of Education Reform, Office of Educational Research and Improvement, U.S. Department of Education.

Koretz, D. M., Barron, S., Mitchell, K. J., & Stecher, B.M. (1996). Perceived effects of the Kentucky instruction results information district. MR-792-PCT/FF . Santa Monica, CA: RAND.

Koretz, D. M., Mitchell, K., Barron, S., & Keith, S. (1996). Final report: Perceived effects of the Maryland School Performance Assessment Program. (CFDA No. 84.117G). National Center for Research on Evaluation, Standards, and Student Testing, LA.

Lane, S., Parke, C.S. & Stone, C.A. (1998). Consequences of the Maryland School Performance Assessment Program. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.

Lane, S., Parke, C.S., & Stone, C. A. (1999, March). MSPAP Impact Study: Mathematics. University of Pittsburgh.

Linn, R. L. (1993). Educational assessment: Expanded expectations and challenges. Educational Evaluation and Policy Analysis, 15(1), 1-16.

Linn, R. L. (1994). Performance assessment: Policy promises and technical measurement standards. Educational Researcher, 23(9), 4-14.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. Educational Researcher, 20(8), 15-21.

Maryland State Board of Education (1995). Maryland school performance report: State and school systems. Baltimore, MD.

McDonnell, L.M. & Choisser, C. (1997, Septmeber). Testing and teaching: Local implementation of new state assessments. CSE Technical Report 442. National Center for Research on Evaluation, Standards, and Student Testing (CRESST) Center for the Study of Evaluation (CSE) Graduate School of Education and Information Studies, University of California, Los Angeles, CA.

Mehrens, W.A. (1998). Consequences of Assessment: What is the Evidence? Evaluation Policy Analysis Archives, 6(13).

Messick, S. (1992). The interplay of evidence and consequences in the validation of performance assessments (ETS RR-92-39). Princeton, NJ: Educational Testing Service.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational Measurement (3rd ed.) (pp. 13-104). New York: American Council on Education.

National Council on Education Standards and Testing. (1992). Raising standards for American education. Washington, DC: Author.

Pomplum, M. (1997). State assessment and instructional change: A path model analysis. Applied Measurement in Education, 10(3), 217-234.